Report No. 66-57B
April 1967

IDENTIFICATION OF LINEAR STOCHASTIC DYNAMICAL SYSTEMS BY NONLINEAR WEIGHTING OF NOISY MEASUREMENTS

Walter Henry Kroy, Jr.

Department of Engineering
University of California
Los Angeles, California

FOREWORD

The research described in this report, "Identification of Linear Stochastic Dynamical Systems by Nonlinear Weighting of Noisy Measurements," Number 66-57B, by Walter Henry Kroy, Jr., was carried out under the direction of C. T. Leondes, E. B. Stear, and A. R. Stubberud, in the Department of Engineering, University of California, Los Angeles.

This project is sponsored in part by the National Aeronautics and Space Administration under Grant NsG-237-62 to the Institute of Geophysics and Planetary Physics of the University.

This report was the basis for a dissertation submitted by the Author.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF SYMBOLS

		Page
a	scalar plant parameter	4
$\bar{a}$	mean of scalar parameter	4
a_{ij}	i, j^{th} element of A	1
$\hat{a}_1(k)$	minimum mean square error linear estimate of a, based upon observations up to the k^{th} . .	12
$\hat{a}_2(k)$	minimum mean square error quadratic estimate of the random parameter a, based upon observations up to time k	14
A	plant matrix: matrix of coefficients associated with dynamical system	1
$\underline{b}$	vector constructed from those elements of the plant matrix which are random	61
$c_1(j)$	$E\, a\, \Delta\, Y_1(j)$	13
$c_2(j)$	$E\, a\, \Delta\, Y_2(j)$	16
c_n	see Section 6.6	57
$D(a;S)$	Fredholm Determinant $= \lim_{n \to \infty} D_n(a;S)$	42
$D_n(a;S)$	see Proof, Theorem 4.3	40
$\overline{e^2}$	mean square error	1
$\overline{e_m^2(n)}$	the mean square error between the plant parameter and its m^{th} degree polynomial estimate at the sample time n	20
E	expectation operator	1
$\hat{E}$	minimum variance linear estimate.	16
$\hat{f}(a)$	Bayes estimate of $f(a)$ given $z(t)$ over $[0, T]$	45
$g(a)$	see Section 4.3	34
$g_1(t_1, t_2)$	see Theorem 6.1	50
$g_2(t_1, t_2; t_3, t_4)$	see Theorem 6.1	50
$g(t_1, t_2; a)$	see Proof of Theorem 4.3	41

		Page
$g_n(t_1, t_2; a)$	see Proof of Theorem 4.3	41
G	operator, see Remarks, Section 6.5	55
G_1	maximum of $g(t_1, t_2)$ see Theorem 6.2	53
G_2	maximum of $g(t_1, t_2, t_3, t_4)$ see Theorem 6.2	53
$h(a)$	see Section 4.3	34
$h_n(a)$	see Equation (4.26)	30
$I(a)$	$\lim_{n \to \infty} I_n(a)$	56
$I_n(a)$	information in the sample; see Section 6.6	50
$k_1(a)$	coefficient of $\phi_1(t;a)$	27
$m_1(i, j)$	$E\ y(i)\ \Delta\ y(j)$	13, 71
$m_2(k-j, k)$	$E\ \Delta\ Y_2(k-j)\ Y_2^*(k)$	16
$m_2^*(k, k-j)$	$E\ Y_2(k)\ \Delta\ Y_2^*(k-j)$	16
$m_{2o}(t_1, t_2)$	$E(a-\bar{a})\ z(t_1)\ z(t_2)$	50
N_o	covariance of η, i.e., $E\ \eta_i\ \eta_j = N_o\ \delta_{ij}$	1
$p(a)$, $p_A(a)$	probability density of the plant parameter a	8
$p_1(j)$	$m_1(j, j)$	13
$p_A(a \mid z(t),\ 0 \le t \le T)$	conditional probability density of a, given the realization $z(t)$ over the time interval 0 to T	45
$p_{Z\mid A}(Z \mid a)$	conditional probability of the vector Z when the plant parameter a is available	10
$p_{Z\mid A}(z_o, z_1, \ldots, z_k \mid a)$	conditional probability ot the observations $z_o, \ldots, z_k$ given a	9

LIST OF SYMBOLS (Continued)

LIST OF SYMBOLS (Continued)

		Page
$W^{(1)}(t)$	linear weighting function at time t	46
$W_n(t_1, t_2)$	n^{th} successive iteration	53
$W(t_1, t_2)$ $W^{(2)}(t_1, t_2)$	quadratic weighting function at time t_1 and t_2	52
x	state of dynamical system	1
x_k, $x(k)$	state at time k	1
x_o or $x(o)$	state of dynamical system at time zero	1
$y(j)$	observation at time j about its mean: $y(j) = z(j) - z(j)$	1
$y(t_1, t_2)$	$y \in R$	52
$Y_1(k)$	observation vector about its mean $Y_1(k) = Z_1(k) - E\ Z_1(k)$	17
$\mathcal{Y}_2(k)$	$\mathcal{Z}_2(k)$ about its mean	15
z	observation or measurement of state x in additive noise η	1
z_k, $z(k)$	observation at time k or k^{th} eigenvalue associated with eigenfunction expansion of $z(t)$	1, 29
z_i	i^{th} eigenvalue of $z(t)$ where $i=1, 2, \ldots$	29
$\underline{Z}(t)$	measurement or observation vector.	61
$Z_1(k)$	observation vector consisting of the previous k observation $Z_1(k) = [z(o), z(1), \ldots, z(k)]^*$ ...	12
$Z_m(k)$	see Section 3.3	21

LIST OF SYMBOLS (Continued)

		Page
$Z_2(k)$	vector consisting of the k^{th} observation and the products formed by it and $z(i)$, $i=0,1,\ldots,k$. That is, $Z_2(k) = \left[z(k), z(k)z(o), z(k)z(1), \ldots, z(k)^2\right]^*$	15
Z_k	vector of observations $z_o, \ldots, z_k$	10
$Z_3(k)$	see Section 3.3	19
$Z_1(k)$	vector of observations $z_o, \ldots, z_k$ same as Z_k	12
$\mathfrak{Z}_2(k)$	vector of observations $\left[Z_o^*, Z_1^*, \ldots, Z_k^*\right]^*$ where $Z_k^* = \left[z_k, z_k z_o, z_k z_1, z_k z_2, \ldots, z_k^2\right]^*$ . . .	15
$\mathfrak{Z}_m(k)$	vector consisting of the k^{th} observation and certain products formed by polynomials of degree m. That is, $Z_m(k) =$ $[z(k), z(k)z(o), \ldots, z(k)^2, z(k)z(o)z(o), \ldots, z(k)^m]$	20
$\mathfrak{Z}^{(2)}(t_1, t_2)$	4 element vector generated by products of the elements from $\underline{Z}(t_1)$ and $Z(t_2)$	61
$\mathfrak{Z}_2(k)$	vector formed by the direct sum of $Z_2(0), Z_2(1), \ldots, Z_2(k)$ and containing α elements where $\alpha = \frac{(k+1)(k+4)}{2}$	15
α	scalar	15
$\underline{\alpha}$	vector whose elements are the elements of the A matrix arranged as a vector $[a_{11}, a_{12}, \ldots, a_{1n}, a_{21}, a_{22}, a_{23}, \ldots, a_{nn}]^*$	1
$\alpha_m(k)$	see Section 3.3	20
$\alpha(a_o, a_1, t)$	coefficient in the expansion of e^{At}	60
$\beta(a_o, a_1, t)$	coefficient in the expansion of e^{At}	60

LIST OF SYMBOLS (Continued)

		Page
$\Lambda_k(a)$	covariance of $\mathfrak{z}_k$ with elements λ_{ij}, x_o known	8
$\mu_i(a)$	eigenvalue associated with $R_{Z\|A}(t_1, t_2)$ see Equation (4.11)	27
ξ	white gaussian disturbing noise acting on the dynamical system	1
$\xi_k, \xi(k)$	disturbing noise at time k	1
$\rho^2(x, y)$	metric on R	52
ρ^2_{n+1}	$\rho^2(W_{n+1}, W_n)$ distance squared between W_{n+1} and W_n in R	54
σ	variance of scalar random plant parameter a .	1
$\phi_i(t;a)$	eigenfunction associated with $R_{Z\|A}(t_1, t_2)$	27
$\psi_i(t;a)$	orthogonal set of eigenfunctions	28
$\|\cdot\|$	determinant or absolute value of quantity in between the two vertical lines	10,11

<u>SUBSCRIPTS</u>

B	denotes estimate is a Bayes estimate
i	index
j	index
k	index
ℓ	index
m	degree of polynomial considered
M	denotes m^{th} degree polynomial considered
n	number of samples
o	denotes the index is zero, hence the initial condition

SUBSCRIPTS (Continued)

r	number of elements in the vector $\underline{x}$
t	time
α	index
β	index
-	underline under a symbol denotes the symbol represents a vector, hence an array

SUPERSCRIPTS

$\wedge$	circumflex denotes the estimate of the variable it is over
*	denotes the transpose of the matrix or vector it is associated with
-	bar; denotes expected value of the variable it is over; expectation may be conditional expectation
-1	inverse of matrix or scales
.	$\frac{d}{dt}$

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Statement of the Problem for DISCRETE Time Parameterization

Consider the vector linear stochastic dynamical system given by the equation

$$\underline{x}_{k+1} = A\,\underline{x}_k + \underline{\xi}_k \qquad (1.1)$$

where $\underline{x}_k$ is an r element vector (i.e., an $r \times 1$ matrix) representing the state of the dynamical system at time k. x_o is the random initial condition assumed here to be a vector gaussianly distributed with mean $\bar{x}_o$ and covariances matrix $E\,x_o x_o^* = P_o$. A is an $r \times r$ matrix of gaussian random parameters a_{ij} with known mean and covariance elements $E(a_{i_1 j_1} - E\,a_{i_1 j_1})(a_{i_2 j_2} - E\,a_{i_2 j_2})$ for $i_1, i_2, j_1, j_2 = 1, 2, 3, \ldots, r$. ξ_k is an r element vector gaussian white noise process with zero mean and covariance function $E\,\xi_k \xi_\ell = R_o\,\delta_{k\ell}$ where $R_o \geq 0$ and δ_{ij} is the Kronecker delta function.

Consider the above dynamical system observed in additive noise. Let z_k be the r dimensional observation vector of the state x_k and η_k the gaussian white noise sequence with zero mean and covariance $E\,\eta_k \eta_\ell = N_o\,\delta_{k\ell}$ where $N_o > 0$ and δ_{ij} is the Kronecker delta function. η_k, ξ_ℓ, a_{ij} and x_o are assumed statistically independent of each other.

The problem is, given the measurements $z_o, z_1, \ldots, z_k$, estimate the random plant parameters a_{ij} in such a way that the weighted mean square error

$$\overline{e^2} = E(\underline{\alpha} - \underline{\hat{\alpha}})^* W(\underline{\alpha} - \underline{\hat{\alpha}}) = \text{trace } E(\underline{\alpha} - \underline{\hat{\alpha}})(\underline{\alpha} - \underline{\hat{\alpha}})^* W^*$$

is minimized. The asterisk above denotes the matrix (vector) is to be transposed. $W = [w_{k\ell}]$ is an $r^2 \times r^2$ positive definite matrix of weighting elements with $k, \ell = 1, 2, \ldots, r^2$, while

$$\underline{\alpha} = \left[\alpha_1, \alpha_2, \ldots, \alpha_{r^2}\right]^* \quad \text{where } \alpha_\ell = a_{ij}, \ \ell = r(i-1) + j,$$

and $i, j = 1, 2, \ldots, r$. $\hat{\alpha}$ is the estimate of $\underline{\alpha}$ and hence of A.

The optimum minimum mean square estimate $\hat{\alpha}_B(k)$ of α given $z_o, z_1, \ldots, z_k$ is the conditional expectation $E[\alpha | z_o, z_1, \ldots z_k]$ or Bayes estimate.

The Bayes estimate is the nonlinear functional that is to be investigated. The scalar case, $r = 1$, will be investigated in detail, and then extensions to the matrix case developed.

1.2 Statement of the Problem for CONTINUOUS Time Parameterization

Consider now the vector linear stochastic dynamical system given by the equation:†

$$\frac{d\,\underline{x}(t)}{dt} = A\,\underline{x}(t) + \underline{\xi}(t) \qquad 0 \le t \le T$$

where $\underline{x}(t)$ is an r element vector function of time t, $\underline{x}(o) \equiv \underline{x_o}$ is the initial state of the system and assumed here to be a gaussian

---

†The differential equation should actually have been written in the differential form $d\,x_t = A\,x_t\,dt + R_t^{\frac{1}{2}}\,d\beta_t$ of References 6, 20, and 21, where β_t is a vector process of independent Brownian motions and x_t for known A, a Markov process independent of the β_t process, R_t is positive definite, so that heuristically, $\xi(t)$ may be thought of as $R_t^{\frac{1}{2}} \frac{d\beta_t}{dt}$ when R_t is a constant. Since $\xi(t)$ in the end will always be represented in its integral form, use of the Stieltjes[30] form, as used in the references, appears cumbersome, so that it will not be used here, but understood to be the underlying structure.

vector random parameter with mean $\overline{x_o}$ and covariance matrix $E\, x_o x_o^* = P_o$. As above, A is an r x r matrix of gaussian random parameters with known mean and covariance elements,

$$E\left(a_{i_1j_1} - E\, a_{i_1j_1}\right)\left(a_{i_2j_2} - E\, a_{i_2j_2}\right)$$

where $i_1, i_2, j_1, j_2 = 1, 2, \ldots, r$. $\xi(t)$ is an r element vector gaussian white noise process with zero mean and covariance function,

$$E\, \xi(t_1)\, \xi(t_2) = R_o\, \delta(t_1 - t_2),$$

where $R_o \geq 0$. $\delta(t_1 - t_2)$ is the so-called impulse function.

Consider the dynamical system to be observed in additive noise; that is,

$$z(t) = x(t) + \eta(t) \qquad\qquad 0 \leq t \leq T$$

where x(t) is the state at time t and $\eta(t)$ is gaussian additive white noise of zero mean and its covariance is given by

$$E\, \eta(t_1)\, \eta(t_2) = N_o\, \delta(t_1 - t_2).$$

The problem now is, given the measurements z(t), $0 \leq t \leq T$, estimate the random plant parameter a_{ij} in such a way that the weighted mean square error

$$\overline{e^2} = E(\underline{\alpha} - \hat{\underline{\alpha}})^*\, W(\underline{\alpha} - \hat{\underline{\alpha}})$$

is minimized. The asterisk denotes transposition of the associated matrix. $W = [w_{k\ell}]$ is an $r^2 \times r^2$ positive definite matrix of weighting elements with $k, \ell = 1, 2, \ldots, r^2$. As before,

$$\underline{\alpha} = \left[\alpha_1, \alpha_2, \ldots \alpha_{r^2}\right]^* .$$

where $\alpha_\ell = a_{ij}$, $\ell = r(i-1) + j$ and $i, j = 1, 2, \ldots, r$. $\hat{\underline{\alpha}}$ is the estimate of $\underline{\alpha}$ and hence of A.

The optimum minimum mean square estimate $\hat{\alpha}_B(k)$ of α when $z(t)$, $0 \le t \le T$, is available is the conditional expectation

$$\underline{\hat{\alpha}}_B(k) = E\left\{\underline{\alpha} \middle| z(t),\ 0 \le t \le T\right\} .$$

The conditional expectation of α given the realization $z(t)$, $t \in [0, T]$ is known as the Bayes estimate[14] when the error criterion is the minimum mean square error criterion. One should note that $z(t)$ involves $x(t)$, and $x(t)$ in turn involves $e^{At} x_o$. e^{At} is not gaussian. This is true even in scalar case where $A = a$. For fixed t, e^{at} is lognormally distributed[2] and so, when the gaussian x_o is multiplied by it, $e^{at} x_o$ is for fixed t non-gaussianly distributed. Consequently $z(t)$ is non-gaussianly distributed. Intuitively one would then suspect nonlinear weighting of $z(t)$, $t \in [0, T]$ improves the estimate over a linear estimate. In fact, if $\overline{x_o} = 0$, the minimum variance linear estimate of a is $\bar{a}$, the mean of a_j. Observing $z(t)$ gives no new information on a.

The approach to be followed in the succeeding chapters is to first consider the scalar cases, i.e., the scalar discrete case and then the scalar continuous case. Certain computation difficulties will be seen to arise, which can be overcome by using polynomial estimators. Last, the vector dynamical case will be treated.

1.3 Examples of Physical Systems

Linear dynamical systems of the form $\frac{d\,x(t)}{dt} = A\,x(t) + \xi(t)$ or $x_{k+1} = A\,x_k + \xi_k$ arise in many engineering problems, e.g., in RCL circuits with thermal noise, lumped parameter mechanical systems subjected to random vibrations, and guidance of space vehicles, just to name a few. In each of these examples, the A matrix may not have been known other than with respect to some nominal value, or it might have been known only with respect to a set of means and variances. It is desired to estimate A based upon the output of the dynamical system.

Another class of problems occurs whenever the A matrix changes from its original value. Such a situation occurs whenever components deteriorate, either from a harsh environment or from age. In such cases, a new estimate of the plant is desired.

In certain space guidance applications, the linearized equations of motion describing small variantions about the nominal trajectory depend upon certain parameters, such as mass, inertia, configuration, rigidity, etc. These quantities are known only approximately. By re-identifying the plant, a more accurate estimate of the system is obtained. In each of the above examples, the systems are to be identified even if the measurements are noisy.

1.4 Historical Background

Many authors have investigated the identification of linear stochastic systems by noisy measurements. They have suggested to increase the dimension of the state space and either

1) do essentially piecewise linear operations, recalculating new gains at each step, such as described by Gunkel,[17] Lee,[27] Stubberud,[39] Ohap,[34] Seal,[37] or Kopp and Orford,[24] or
2) determine certain gradients, and then solve the resulting differential equation by using the quasi-linearization of Kalaba, as described by Kumar and Sridhar,[25] or
3) expand the function describing the nonlinear plant in terms of a taylor series, retaining terms up to second order in the state variable, do the same

for the observation function and then solve the system of equations for the best estimate as suggested by Bass, et al.[5]

The optimum minimum mean square estimate is still the Bayes or conditional expectation of the modified state given the observables. The above are approximations to the Bayes estimate. None appear to have given data on how well they approximate the Bayes estimate, particularly if the statistics are strongly non-gaussian and/or the noise levels are large. In each of the above cases, the processing of the data does not occur until after the measurements have been made. This is not the case however, when the polynomial estimators are of the form suggested by Cameron and Martin,[7] Masani and Wiener,[37] A.V. Balakrishnan,[3,4] or as developed herein. As pointed out by Balakrishnan,[3] Polya and Szego[41] have shown that it is possible to construct processes for which one cannot produce a convergent polynomial sequence. It is possible, however, to construct an expansion in terms of nonlinear functionals of the observables which will give point-wise convergence to the Bayes estimate.[3]

Another approach is to solve a partial differential equation (PDE) in the form of the Fokker-Planck or Komogorov diffusion equation for the conditional probability density. This was suggested by Stratonovitch,[38] Kushner,[26] Fisher,[13] and Mortensen.[33] The technique of Stratonovitch (corrected by Kushner, which in turn was corrected by Fisher) requires solving a set of nonlinear partial differential equations which are coupled. Fisher has in addition developed a partial differential equation, which, when solved, gives the conditional mean and a limited number of quasi-moments. In any case, the partial differential equations are highly coupled in the random parameters. Consequently, they do not lend themselves to solution, but only to approximate solutions. Mortensen[38] has developed a PDE which, when normalized, becomes the conditional probability density of the amended state space given the observation over the time interval [0, T]. In this way he avoids certain coupling between terms of the PDE associated with the scalar dynamical system where no disturbing noise is acting but there is measurement noise. The PDE he obtains is of first order and can be solved. With disturbing noise, the PDE of Mortensen becomes a second order quasilinear PDE, which at least to date has not been solved.

In short, the identification problem posed in the dissertation has not been investigated by others in the manner proposed herein.

CHAPTER II

BAYES ESTIMATE OF SCALAR PLANT PARAMETER – DISCRETE TIME CASE

2.1 Statement of the Problem

The objective of this chapter is to develop the minimum mean square estimate of the scalar random plant parameter a. The problem is as posed in Section 1.1, namely, given the scalar dynamical system, $x_{k+1} = a\,x_k + \xi_k$, and the observations, $z_k = x_k + \eta_k$, where $k = 0, 1, \ldots, \eta$, estimate the scalar random plant parameter a with known distribution. ξ_k, η_k are statistically independent white gaussian sequences with zero means such that $E\,\xi_i\xi_j = R_o\,\delta_{ij}$, $E\,\xi_i\eta_j = 0$, $E\,\eta_i\eta_j = N_o\,\delta_{ij}$. x_o is the initial state of the system.

2.2 The Initial State x_o is known

The Bayes estimate of a, based upon the k+1 observations $z_o, z_1, \ldots, z_k$ is

$$\hat{a}_B(k) = E[a|z_o \ldots z_k] = \frac{\int a\,p(a)\,p(z_o, \ldots, z_k|a)\,da}{\int p(a)\,p(z_{o1}, \ldots, z_k|a)\,da} \tag{2.1}$$

$p_{Z|A}(z_o, \ldots, z_k|a)$ is a gaussian conditional probability density. Denote the observations as a vector $Z_k = [z_o, z_1, \ldots, z_k]^*$, and its mean conditioned on a by Z_k. Let

$$\Lambda_k(a) \equiv E\Big((Z_k - Z_k)(Z_k - Z_k)^*\,|a\Big) \equiv \begin{bmatrix} \lambda_{oo} & \lambda_{o1} & \cdots & \lambda_{ok} \\ \lambda_{1o} & \lambda_{11} & \cdots & \lambda_{1k} \\ \vdots & \vdots & & \vdots \\ \lambda_{ko} & \lambda_{k1} & & \lambda_{kk} \end{bmatrix} \tag{2.2}$$

where

$$\lambda_{oj} = \lambda_{jo} = N_o\, \delta_{oj} \qquad j = 0, 1, \ldots, k$$

$$\lambda_{ij} = \lambda_{ji} = R_o \sum_{\alpha=1}^{i} a^{i+j-2} + N_o\, \delta_{ij} \qquad i, j = 1, 2, \ldots, k \tag{2.3}$$

clearly $|\Lambda_k(a)|$ is a polynomial in a for $k \geq 2$. The conditional probability density $p(z_o, \ldots, z_k | a)$ or $p_{Z|A}(Z_k|a)$ becomes

$$p_{Z|A}(Z_k|a) = \frac{e^{-\frac{1}{2}(Z_k - \overline{Z}_k)^* \Lambda^{-1}(a)\,(Z_k - \overline{Z}_k)}}{(2\pi)^{\frac{k+1}{2}}\, |\Lambda(a)|} \tag{2.4}$$

For $k=0$ and for $k=1$, $p_{Z|A}(Z|a)$ is independent of a, consequently the Bayes estimate for the first two samples can be readily determined. In the case where a is gaussianly distributed with mean $\bar{a}$ and variance σ, the Bayes estimate becomes

$$\hat{a}_B(0) = \bar{a} \qquad\qquad x_o \text{ Known} \tag{2.5}$$

$$\hat{a}_B(1) = \frac{\dfrac{\bar{a}}{\sigma^2} + \dfrac{x_o z_1}{N_o + R_o}}{\dfrac{1}{\sigma^2} + \dfrac{x_o^2}{N_o + R_o}}$$

$\hat{a}_B(1)$ does not depend upon z_o (as should be expected) since z_o does not depend upon a. Furthermore as $\sigma \to \infty$,

$$\hat{a}_B(1) = \frac{z_1}{x_o} = \frac{a\, x_o + \xi_o}{x_o} = a + \frac{\xi_o}{x_o}\,.$$

When $\sigma \to 0$ then $\hat{a}_B(1) = \bar{a}$.

For $k \geq 2$, the denominator of $p_{Z|A}(Z_k|a)$ is a positive definite polynomial in a, while the argument of the exponential is the ratio of two polynomials in a, the numerator of which depends upon the observables. As a result the integrals associated with the Bayes estimate cannot be readily evaluated. Numerical results could be obtained by use of a computer. The case, where x_o is a gaussian random scalar parameter, is treated next.

2.3 The Initial State x_o is a Gaussian Random Parameter

Let x_o be a gaussian random parameter with mean $\bar{x}_o$ and variance V_o. The Bayes estimate then becomes for the $k+1$ observations $z_o, \ldots, z_k$:

$$\hat{a}_B(k) = \frac{\int a\, p_A(a)\, p_{Z|A}(Z_k|a)\, da}{\int p_A(a)\, p_{Z|A}(Z_k|a)\, da} \tag{2.6}$$

where

$$Z_k = \begin{bmatrix} z_o \\ z_1 \\ \vdots \\ z_k \end{bmatrix} \quad \text{and} \quad z_\ell = a^\ell x_o + \sum_{j=1}^{\ell} a^{\ell-j} \xi_{j-1}$$

$$p_{Z|a}(Z_k|a) = (2\pi)^{-\frac{k+1}{2}} |\Gamma_k(a)|^{-1} \exp - \left\{(Z_k - Z_k)^* \Gamma_k^{-1}(a)(Z_k - Z_k)\right\} \tag{2.7}$$

$$\Gamma_k = \begin{vmatrix} \gamma_{oo} & \gamma_{o1} & \cdots & \gamma_{ok} \\ \gamma_{1o} & \gamma_{11} & & \\ \vdots & & & \\ \gamma_{ko} & \cdots & \cdots & \gamma_{kk} \end{vmatrix} \tag{2.8}$$

and

$$\begin{aligned} \gamma_{jo} = \gamma_{oj} &= a^j V_o^2 + N_o \delta_{oj} && j = 0, 1, 2, \ldots, k \\ \gamma_{ji} = \gamma_{ij} &= a^{i+j} V_o^2 + R_o \sum_{\ell=1}^{i} a^{i+j-2\ell} + N_o \delta_{jk} && i, j = 1, 2, \ldots, k\,. \end{aligned} \tag{2.9}$$

Clearly $|\Gamma(a)|$ is a polynomial in a for $k \geq 1$. For $k=0$, $\hat{a}_B(0)=E\,a$ as before. When $k \geq 1$, the argument of the exponential associated with $p_{Z|A}(Z|a)$ is a ratio of polynomials in a. When $p_{Z|A}(Z|a)$ is substituted into Equation (2.6) an expression is again obtained, which is not readily evaluated, except by numerical techniques and these in turn depend upon obtaining first a specific realization $(z_o, \ldots, z_k)$.

It is desirable at this point to investigate the feasibility of using estimators based upon polynomial weighting of the observables, the obvious advantage being the weightings do not depend upon the specific realizations, but only upon the known or assumed prior statistics.

CHAPTER III

SCALAR PLANT PARAMETER ESTIMATED BY RECURSIVE POLYNOMIAL WEIGHTING AT DISCRETE TIMES

3.1 Introduction and Linear Estimator

The problem is as stated in Chapter II, namely, let the scalar dynamical system be characterized by

$$x(k+1) = a\, x(k) + \xi(k) \tag{3.1}$$

and the measurements by

$$z(k) = x(k) + \eta(k) \,. \tag{3.2}$$

$x(o)$ is a scalar gaussian initial state with mean $\bar{x}_o$ and variance V_o. a is characterized by $p(a)$. Let $\bar{a} = E\, a$. The other statistics are as before.

Let the first $k+1$ observations be denoted by the vector $Z_1(k) = [z(o), \ldots, z(k)]^*$ and its difference from the mean as $Y_1(k) = Z_1(k) - E\, Z_1(k)$. The optimum linear weighting minimum mean square error estimate $\hat{a}_1(k)$ of a when $k+1$ measurements have been made, is known to be

$$\hat{a}_1(k) = \bar{a} + \left(E\, a\, Y(k)\right) [E\, Y_1(k)\, Y_1^*(k)]^{-1}\, Y_1(k) \tag{3.3}$$

while the corresponding error is known to be

$$\overline{e_1^2(k)} = E(a-\bar{a})^2 - \left(E\, a\, Y_1^*(k)\right) \left(E\, Y_1(k)\, Y_1^*(k)\right)^{-1} E\left(a\, Y_1(k)\right) \tag{3.4}$$

In the case where the mean of the initial state is zero $E\, Y_1(k) = 0$, $E\, a\, Y_1(k) = 0$, and $\overline{e^2(k)} = E(a-\bar{a})^2$, even if the variance of the initial state is large, so that linear weighting does not improve the estimate at all!

Equations (3.3) and (3.4) are readily cast into the set of recursion relations:† The initial conditions are

$$y(j) = z(j) - \overline{z(j)} \qquad j = 0, 1, \ldots, n$$

$$m_1(o, j) = \overline{a^j}\, V_o^2 \qquad j = 0, 1, \ldots, n$$

$$c_1(o) = 0 \tag{3.5}$$

$$\hat{a}_1(o) = E\, a$$

$$\overline{e^2(o)} = E(a-\bar{a})^2 = \text{variance of } a$$

The recursion relations for the coefficients when $j = 1, 2, \ldots, k$ and $k = 1, 2, \ldots, n$ are

$$m_1(k, k-j) = 0$$

$$m_1(k-j, k) = E\, y(k-j)\, y(k) - \sum_{\ell=0}^{k-j-1} m_1^*(k-j, k-j-1-\ell)\, p_1^{-1}(k-j-1-\ell)\, m_1(k-j-1-\ell, k) \tag{3.6}$$

$$p_1(j) = m_1(j, j) \tag{3.7}$$

$$c_1(k) = E\, a\, y(k) - \sum_{\ell=1}^{k} c_1(k-\ell)\, p_1^{-1}(k-\ell)\, m_1(k-\ell, k) \tag{3.8}$$

Thus

$$\Delta\, y(k+1) = y(k+1) - \sum_{\ell=1}^{k} \left[m_1(k+1, k-\ell)\, p_1^{-1}(k-\ell) \right] \Delta\, y(k-\ell) \tag{3.9}$$

$$\hat{a}_1(k+1) = \hat{a}_1(k) + \left[c_1^*(k+1)\, p_1^{-1}(k+1) \right] y(k+1) \tag{3.10}$$

and

$$\overline{e_1^2(k+1)} = \overline{e_1^2(k)} - c_1^*(k+1)\, p_1^{-1}(k+1)\, c_1(k+1) \tag{3.11}$$

Instead of inverting a matrix of $(k+1) \times (k+1)$ dimensions, the recursive linear scheme has led to inversion of only a scalar. Note

† The details of the derivation are primarily algebraic, hence for convenience, they are developed in the appendix, see Appendix A.

that all the weightings are linear. The weightings are determined strictly from the prior statistics. Consider now quadratic weighting, i.e., weighting on polynomials of first and second degree.

3.2 Recursive Quadratic Estimation

Consider the same dynamical system, $x(k+1) = ax(k) + \xi(k)$ and the same measurement system, $z(k) = x(k) + \eta(k)$, as discussed in Section 3.1, Equations (3.1) and (3.2). But now form the vector $Z_2(o) = \begin{bmatrix} z(o) \\ z(o)^2 \end{bmatrix}$ after the observation $z(o)$ at time 0 becomes available. Then consider the vector $Z_2(1) = \begin{bmatrix} z(1) \\ z(1)\, z(o) \\ z(1)^2 \end{bmatrix}$ formed from $z(o)$ and $z(1)$. The k^{th} observation $z(k)$ allows the vector

$$Z_2(k) = \begin{bmatrix} z(k) \\ z(k)\, z(o) \\ z(k)\, z(1) \\ \vdots \\ z(k)^2 \end{bmatrix}$$

to be formed. Let $a_2(k)$ be the estimate obtained by an optimum linear weighting upon the array $\begin{bmatrix} Z_2(o) \\ Z_2(1) \\ \vdots \\ Z_2(k) \end{bmatrix}$.

This array has $\frac{(k+1)(k+4)}{2}$ elements. Optimum linear weighting upon an array with these elements is equivalent to forming the optimum weights, W_o, W_1^i, W_2^{ij} for the polynomial

$$W_o + \sum_{i=0}^{k} W_1^i\, z(i) + \sum_{i=0}^{k} \sum_{j=0}^{k} W_2^{ij}\, z(i)\, z(j) \;.$$

Because terms which are up to and including second order in the observables are considered, the scheme proposed has been called "quadratic estimation". Polynomial estimators, which weigh observables up to third degree, are called cubic estimators, etc.

The minimum mean square error quadratic estimate of the plant parameter a, based upon the vectors $Z_o, \ldots, Z_k$, is obtained as follows: Let

$$Z_2(k) = \begin{bmatrix} z(k) \\ z(k)\ z(o) \\ z(k)\ z(1) \\ \vdots \\ z(k)\ z(k) \end{bmatrix}_{(k+2)x1} \quad \text{and} \quad \mathcal{Z}_2(k) = \begin{bmatrix} Z_2(o) \\ Z_2(1) \\ \vdots \\ Z_2(k-1) \\ Z_2(k) \end{bmatrix}_{\alpha x1} = \begin{bmatrix} \mathcal{Z}_2(k-1) \\ \cdots \\ Z_2(k) \end{bmatrix}_{\alpha x1} \tag{3.12}$$

where $\alpha = \frac{(k+1)(k+4)}{2}$. Let

$$Y_2(k) = Z_2(k) - E\, Z_2(k) \quad \text{and} \quad \mathcal{Y}_2(k) \equiv \mathcal{Z}_2(k) - E\, \mathcal{Z}_2(k) \tag{3.13}$$

Let $\Delta Y_2(k)$ be defined as the difference between $Y_2(k)$ and the minimum mean square linear estimate of $Y_2(k)$ given $Z_2(o), Z_2(1), \ldots, Z_2(k-1)$ or equivalently, given $\Delta Y_2(o), \ldots, \Delta Y(k-1)$; then

$$\Delta Y_2(k) = Y_2(k) - \left[E\, Y_2(k)\, \Delta \mathcal{Y}_2(k-1)\right]\left[E\, \Delta \mathcal{Y}(k-1)\, \Delta \mathcal{Y}^*(k-1)\right]^{-1} \Delta \mathcal{Y}_2(k-1) \tag{3.14}$$

$E\, \Delta \mathcal{Y}_2(k)\, \Delta \mathcal{Y}_2^*(k)$ has by construction the property

$E\, \Delta \mathcal{Y}(k)\, \Delta \mathcal{Y}^*(k) =$

$$= \begin{bmatrix} E\Delta Y_2(o)\Delta Y_2^*(o) & 0 & 0 & \cdot\ \cdot & 0 \\ 0 & E\Delta Y_2(1)\Delta Y_2^*(1) & 0 & \cdot\ \cdot & 0 \\ 0 & 0 & E\Delta Y_2(2)\Delta Y_2^*(2) & \cdot\ \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot\ \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot\ \cdot & \cdot \\ 0 & 0 & 0 & \cdot\ \cdot & E\Delta Y_2(k)\Delta Y_2^*(k) \end{bmatrix} \tag{3.15}$$

Hence

$$\Delta Y_2(k)=Y_2(k)-\sum_{j=1}^{k}\left[E\,Y_2(k)\,\Delta\,Y_2^*(k-j)\right]\left[E\,\Delta\,Y_2(k-j)\,\Delta\,Y_2^*(k-j)\right]^{-1}\Delta Y_2(k-j) \tag{3.16}$$

where $\Delta Y_2(k)$ is a $(k+1) \times 1$ matrix, i.e., a $(k+1)$ element vector.

Define

$$m_2(k-j,k) = E\,\Delta\,Y_2(k-j)\,Y_2^*(k) \tag{3.17}$$

$$m_2^*(k,k-j) \equiv E\,Y_2(k)\,\Delta\,Y_2^*(k-j) \tag{3.18}$$

$$p_2(j) = E\,\Delta\,Y_2(j)\,\Delta\,Y_2^*(j) \tag{3.19}$$

$$c_2(j) = E\,a\,\Delta\,Y_2(j) \tag{3.20}$$

Substituting Equation (3.16) into Equations (3.17) and (3.20) yields

$$m_2(k-j,k) = E\,Y(k-j)\,Y^*(k)-\sum_{\ell=1}^{k-j-1} m_2(k-j,k-j-1-\ell)p_2^{-1}(k-j-1-\ell)m_2(k-j-1-\ell,k) \tag{3.21}$$

and

$$c_2(k) = E\,A\,Y(k) - \sum_{\ell=1}^{k} m_2^*(k,k-\ell)\,p_2^{-1}(k-\ell)\,c_2(k) \tag{3.22}$$

The optimum quadratic estimate of a given $z(o), \ldots, z(k)$ is of course the optimum linear estimate of a given $Y_2(o), Y_2(1), \ldots, Y_2(k)$. But this estimate of a is also equivalent to the optimum linear estimate given $Y_2(o), Y_2(1), \ldots, Y_2(k)$; that is, if $\hat{E}$ denotes the best <u>linear</u> estimate of the arguments (Doob, Reference 11) then it follows

$$\begin{aligned}\hat{a}_2(k) &= \hat{E}\left(a \,\middle|\, Z_2(o),\; Z_2(1), \ldots, Z_2(k)\right)\\ &= \hat{E}\left(a \,\middle|\, Y_2(o),\; Y_2(1), \ldots, Y_2(k)\right)\\ &= \hat{E}\left(a \,\middle|\, \Delta\,Y_2(o),\; \Delta\,Y_2(1), \ldots, \Delta\,Y_2(k)\right) = E\left(a \,\middle|\, \Delta\mathcal{Y}_2(k)\right)\end{aligned} \tag{3.23}$$

which is

$$\hat{a}_2(k) = \bar{a} + \left(E\, a\, \Delta \mathcal{Y}_2(k)\right)\left[E\, \Delta \mathcal{Y}_2(k)\, \Delta\, \mathcal{Y}_2(k)\right]^{-1} \Delta \mathcal{Y}(k) \qquad (3.24)$$

and this can be reduced because $\Delta \mathcal{Y}_2(k)$ has the property indicated in Equation (3.15) to

$$\hat{a}_2(k) = \hat{a}_2(k-1) + [E\, a\, \Delta\, Y_2^*(k)]\, [E\, \Delta\, Y_2(k)\, \Delta\, Y_2^*(k)]^{-1} \Delta\, Y_2(k)$$

Use of Equations (3.17) through (3.20) results then in $\hat{a}_2(k)$ having the expression

$$\hat{a}_2(k) = \hat{a}_2(k-1) + c_2(k)\, p_2^{-1}(k)\, \Delta\, Y(k)$$

The mean square error in estimating a by the quadratic estimate $\hat{a}_2(k)$ is

$$\overline{e_2^2(k)} = E\, [a-\hat{a}_2(k)]^2 \qquad (3.25)$$

which in turn reduces from

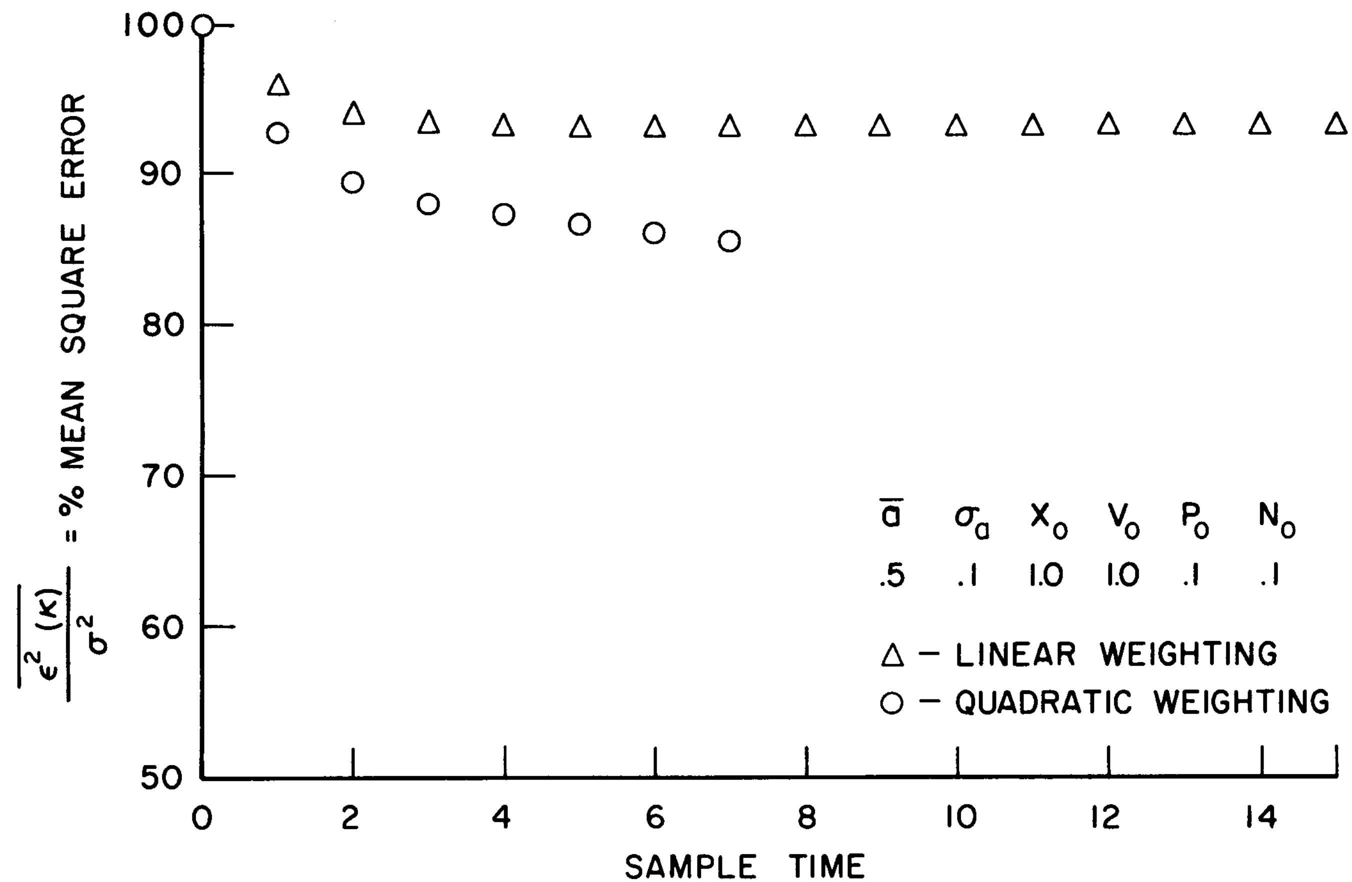
$$\overline{e_2^2(k)} = E(a-\bar{a})^2 - [E\, a\, \Delta \mathcal{Y}_2(k)]\, [E\, \Delta \mathcal{Y}_2(k)\, \Delta \mathcal{Y}_2(k)]^{-1}\, [E\, a\, \Delta \mathcal{Y}_2(k)]$$

to

$$\overline{e_2^2(k)} = \overline{e_2^2(k-1)} - c_2^*(k)\, p_2^{-1}(k)\, c_2(k) \qquad (3.26)$$

Equations (3.24) and (3.26) give the desired recursion relations. The coefficients are given by Equations (3.21) and (3.22).

An example, showing the above technique has been programmed for the IBM 7094 in FORTRAN IV language and is discussed in Appendix B. a is considered gaussian with mean $\bar{a}$ and variance σ. The reduction in mean square error between the linear and quadratic estimators is shown in Figure 1.



PERCENT MEAN SQUARE ERROR VERSUS TIME

FIGURE 1

3.3 Generalization to Higher Ordered Polynomial Weighting

The previous analysis is readily extended to higher dimensions. With no loss of generality, one may consider the cubic estimator by defining

$$Z_3(k) = \begin{bmatrix} z(k) \\ z(k)\ z(o) \\ z(k)\ z(1) \\ \vdots \\ z(k)\ z(k) \\ z(k)\ z(o)\ z(o) \\ z(k)\ z(1)\ z(o) \\ z(k)\ z(1)\ z(1) \\ z(k)\ z(2)\ z(o) \\ \vdots \\ z(k)\ z(k)\ z(k) \end{bmatrix}_{\gamma_3(k)\mathrm{x}1}$$

where

$$\gamma_3(k) = \text{number of rows in } Z_3(k)$$

$$= \frac{(k+2)(k+3)}{2}$$

The size of the $Z_3(k)$ vector grows as

k	0	1	2	3
$\gamma_3(k)$	3	6	10	15

If only a few observations are made, and then processed, cubic estimators appear feasible. The derivations of Section 3.2 go over directly, i.e., one need only substitute $Z_3(k)$ for $Z_2(k)$ in all the equations.

Similarly, polynomial weighting of m^{th} degree requires only to set $Z_m(k)$ to

$$Z_m(o) = \begin{bmatrix} z(o) \\ z^2(o) \\ z^3(o) \\ z^4(o) \\ \vdots \\ z^m(o) \end{bmatrix}_{m\times 1} \qquad Z_m(1) = \begin{bmatrix} z(1) \\ z(1)\ z(o) \\ z(1)\ z(1) \\ z(1)\ z(o)\ z(o) \\ z(1)\ z(1)\ z(1) \\ \vdots \\ z^m(1) \end{bmatrix}_{\alpha_m(1)\times 1}$$

$$Z_m(=) = \begin{bmatrix} z(2) \\ z(2)\ z(o) \\ z(2)\ z(1) \\ z(2)\ z(2) \\ z(2)\ z(o)\ z(o) \\ z(2)\ z(1)\ z(1) \\ \vdots \\ z^m(2) \end{bmatrix}_{\alpha_m(2)\times 1}$$

The number of rows in the vectors are

$$\alpha_m(1) = \frac{m(m+1)}{2}$$

$$\alpha_m(2) = \frac{m(m+1)\ (m+2)}{6}$$

and so forth. Figure 2 gives the growth rate of $Z_m(k)$. Figure 3 gives the growth rate of the vector $\mathfrak{z}_m(n) = \begin{bmatrix} Z_m(o) \\ \vdots \\ Z_m(n) \end{bmatrix}_{\gamma_m(n)\times 1}$

Use of the recursion relations reduces the size of the matrices appreciably. One needs only to compare the square of the length of m = Order of weighting

n = Number of observations available

k = Measurement index (sampling time): k=0, 1, 2, . . . , n

z(k) = Scalar measurement at k^{th} sampling time

$Z_m(k)$ = Vector used in recursive estimator of the form

$$\left[z(k),\ z(k)\,z(o),\ \ldots,\ z(k)\,z(k),\ z(k)\,z(o)\,z(o),\ z(k)\,z(1)\,z(o),\ \ldots,\ z^m(k)\right]^*$$

$\vec{n}$ / ↓m	0	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	2	3	4	5	6	7	8	9	10	11	12	13	14
3	3	6	10	15	21	28	36	45	55	66	78	91	105
4	4	10	20	35	56	84	120	165	220	286	364	455	560
5	5	15	35	70	126	210	320	495	715	1001	1365	1800	
6	6	21	56	126	352	562	882	1377					
7	7	28	84	210	562	1124	2006						
8	8	36	120	320	882	2006							
9	9	45	165	495	1377								
10	10	55	220	715	2092								
11	11	66	286	1001									
12	12	78	364	1365									
13	13	91	455	1800									

LENGTH OF MEASUREMENT VECTOR $Z_m(n)$

FIGURE 2

m	=	Order of weighting
n	=	Number of observations available
k	=	Measurement index (sampling time). $k=0, 1, \ldots, n$
$z(k)$	=	Scalar measurement at k^{th} sampling time
$Z_m(k)$	=	$\left[z(k), \ldots, z(k)\, z(o), \ldots, z(k)\, z(k), \; z(k)\, z(o)\, z(o), \; z(k)\, z(1)\, z(o), \ldots, z^m(k)\right]^*$
$\mathfrak{Z}_m(n)$	=	$\left[Z_m^*(o), \; Z_m^*(1), \ldots, Z_m^*(n)\right]^*$

$\vec{n}$ / $\downarrow m$	0	1	2	3	4	5	6	7	8	9	10	11	12
1	1	2	3	4	5	6	7	8	9	10	11	12	13
2	2	5	9	14	20	27	35	44	54	65	77	90	104
3	3	9	19	34	55	83	119	164	219	285	363	454	559
4	4	14	34	69	125	209	329	494	714	1000			
5	5	20	55	125	251	461	791	1286					
6	6	27	83	209	461	923	1715						
7	7	35	119	329	791	1715	3431						
8	8	44	164	494	1286								
9	9	54	219	714									
10	10	65	285	1000									
11	11	77	363										
12	12	90	454										
13	13	104	559										

LENGTH OF VECTOR $\mathfrak{Z}_m(n)$

FIGURE 3

the $Z_m(n)$ vector with that of the square of the length of the $\mathfrak{z}_m(n)$ vector for the same n and m. For quadratic weighting, the reduction at the k^{th} sampling is from

$$\left[\frac{(k+1)(k+4)}{2}\right]^2 \quad \text{to} \quad (k+1)^2$$

For cubic and quadratic estimators this reduction is even more pronounced.

High order polynomial weighting appears to be applicable primarily where only a few measurements are available for processing and where it is desirable to extract the maximum amount of information from the information available. Because these optimum weightings depend only upon the statistics and not on the specific realization, the weightings can be determined before the realizations are available. This is particularly convenient in systems where processing of data at the time the realization is made available is at a premium — as is the case in many space vehicles.

In order to fully understand the discrete dynamical system, one should also investigate continuous dynamical systems. In the next two chapters the Bayes and the polynomial estimators are developed for continuous dynamical systems.

CHAPTER IV

BAYES ESTIMATE OF SCALAR PLANT PARAMETER – CONTINUOUS TIME CASE

4.1 Statement of Problem

Consider the class of linear dynamical equations[†] given by

$$\frac{dx(t)}{dt} = a\,x(t) + \xi(t) \qquad\qquad 0 \le t \le T < \infty$$

where $x(t)$ is a scalar function of time, t, $x(o) = x_o$ is the initial value of $x(t)$ and is assumed to be gaussianly distributed with known mean $\bar{x}_o$ and variance V_o. "a" is called the plant parameter with known (or assumed) prior distribution. $\xi(t)$ is a white noise gaussian random process with zero mean and covariance function $E\,\xi(t)\,\xi(\tau) = R_o\,\delta(t-\tau)$. $\xi(t)$, a, x_o are statistically independent of each other. $\delta(t)$ is the so-called impulse delta function. The above system is observed in a noisy environment over an interval of time T, i.e.,

$$z(t) = x(t) + \eta(t)$$

where $\eta(t)$ is a white gaussian noise with zero mean and covariance function $E\,\eta(t_1)\,\eta(t_2) = N_o\,\delta(t_1-t_2)$. $\eta(t)$ is statistically independent of a, x, and $\xi(t)$. The problem is to estimate a, the random plant parameter.

The Bayes estimate, that is, the minimum mean square error estimate, is first developed as a limit of an eigenfunction expansion for the case of no disturbing noise, and then later, with disturbing noise. Theorems 4.1 and 4.2 deal with only measurement noise while Theorems 4.3 and 4.4 deal with both measurement noise and disturbing noise. Theorems 4.1 and 4.3 deal with zero mean initial state. Theorem 4.4 is the general solution to the identification of

---
[†]See Footnote of Page 2.

linear stochastic dynamical systems with independent white gaussian disturbance noise and independent white gaussian measurement noise.

4.2 Theorem 4.1

Given

$$x(t) = x_o e^{at} \text{ or equivalently } \dot{x} = a x, \quad 0 \le t \le T < \infty \qquad (4.1)$$

$$z(t) = x(t) + \eta(t) \qquad 0 \le t \le T < \infty \qquad (4.2)$$

where

x_o is gaussian (0 mean, V_o^2) and is the initial value of $x(t)$, a is a random parameter with probability density $p(a)$ having mean $\bar{a}$.

If

$\eta(t)$ is white gaussian noise with zero mean and $E\,\eta(t)\,\eta(t) = N_o\,\delta(t-\tau)$ where $\delta(t)$ is a Dirac delta function.

Then

the minimum mean squared estimate, the conditional expectation, is

$$\hat{a} = E[a \,|\, z(t),\ 0 \le t \le T] = \frac{\int a\, p(a)\, e^{+\frac{1}{2}Q(a)}\, da}{\int p(a)\, e^{+\frac{1}{2}Q(a)}\, da} \qquad (4.3)$$

where

$$Q(a) = \frac{\frac{1}{N_o}\left(\frac{V_o^2}{N_o}\right)}{1 + \left(\frac{V_o^2}{N_o}\right)\left(\frac{e^{2aT}-1}{2a}\right)} \left[\int_o^T z(t)\, e^{at}\, dt\right]^2 - \log\left[1 + \frac{V_o^2}{N_o}\left(\frac{e^{2aT}-1}{2a}\right)\right] \qquad (4.4)$$

(i) If $T = 0$, then $\hat{a} = E\,a$

(ii) If $0 < |aT| << 1$, $\frac{V_o^2}{N_o} << 1$, and $\frac{V_o^2}{N_o}\left[\int_o^T z(t)\, e^{at}\, dt\right]^2 << 1$

then

$$\hat{a} \approx \bar{a} + \int_0^T \int_0^T z(t_1)\, w(t_1, t_2)\, z(t_1)\, dt,\, dt_2$$

where

$$W(t_1, t_2) = \frac{V_o^2}{2N_o^2}\left(1 - \frac{V_o^2 T}{N_o}\right) E\left\{(a-\bar{a})\, e^{a(t_1+t_2)}\right\}$$

Thus the low "signal-to-noise" $\frac{V_o^2}{N_o} << |$ small sampling-time estimator is a quadratic estimator.

(iii) If in addition to (ii), a is gaussianly distributed with mean $\bar{a}$ and variance σ, then

$$W(t_1, t_2) = W_g(t_1, t_2)$$

$$= \frac{V_o^2 \sigma^2}{2N_o^2}\left(1 - \frac{V_o^2 T}{N_o}\right)(t_1+t_2)\, e^{\bar{a}(t_1+t_2)}\, e^{\frac{\sigma^2}{2}(t_1+t_2)^2}$$

<u>Proof to Theorem 4.1</u>

A. Generation of a complete orthonormal set of functions.

The covariance of $x(t)$ given a is, since $E\, x_o = 0$,

$$R_{x|a}(t_1, t_2) = E\{[x(t_1)\, x(t_2)] \,|\, a\} = V_o^2\, e^{a(t_1+t_2)} . \qquad (4.5)$$

$R_{x|a}(t_1, t_2)$ is clearly symmetric in t_1 and t_2 and

$$\int_0^T \int_0^T R_{x|a}(t_1, t_2)\, dt_1\, dt_2 = V_o^4 \left[\frac{e^{2aT} - 1}{2a}\right] < \infty \qquad (4.6)$$

for all $a < 0$. The covariance of the noise $\eta(t)$ is

$$R_n(t_1, t_2) = E\, \eta(t_1)\, \eta(t_2) = N_o\, \delta(t_1 - t_2) . \qquad (4.7)$$

Proof to Theorem 4.1 (Continued)

$R_n(t_1, t_2)$ is symmetric and $\int_0^T \int_0^T R_n(t_1, t_2)\, dt_1\, dt_2 = NT < \infty$ (4.8)

since $t < \infty$. Thus it follows

$$R_{z|a}(t_1, t_2) = E\{z(t_1)\, z(t_2) | a\} = R_{x|a}(t_1, t_2) + R_n(t_1, t_2) \qquad (4.9)$$

is symmetric and

$$\int_0^T \int_0^T |R_{z|a}(t_1, t_2)|^2\, dt_1\, dt_2 < \infty \qquad (4.10)$$

then† the integral

$$\int_0^T R_{z|a}(t_1, t_2)\, \phi_i(t_2; a)\, dt_2 = \mu_i(a)\, \phi_i(t_1; a) \qquad (4.11)$$

is satisfied for at least one real number $\mu_i \neq 0$ and some function $\phi_i(t;a)$ such that

$$0 < \int_0^T |\phi_i(t;a)|^2\, dt < \infty \qquad (4.12)$$

Expand Equation (4.11), using Equations (4.5), (4.7) and (4.9), then,

$$V_o^2 e^{at_1} \int_0^T e^{at_2}\, \phi_i(t_1; a)\, dt_2 + N_o\, \phi_i(t_1; a) = \mu_i(a)\, \phi_1(t_1; a) \qquad (4.13)$$

Thus we see there exists at least one ϕ_i say ϕ_1, which has the form

$$\phi_1(t_1; a) = k_1(a)\, e^{at_1} \qquad (4.14)$$

but

† See Davenport and Root (Reference 10, p. 373), Courant and Hilbert Reference 8, Chapter II, Arts 4 and 5), or Riesz and Nagy (Reference 36, p. 242).

Proof to Theorem 4.1 (Continued)

$$\int_0^T \phi_1^2(t_1;a)\,dt_1 = 1 = k_1^2 \frac{e^{2aT}-1}{2a}$$

so that

$$\phi_1(t;a) = \left[\frac{2a}{e^{2aT}-1}\right]^{\frac{1}{2}} e^{at} \tag{4.15}$$

If the remaining $\phi_i(t_1, a)$ are chosen orthogonal to $\phi_1(t_1, a)$, they will be such that

$$\int R_{x|a}(t_1, t_2)\, \phi_i(t_2, a)\, da = 0 \qquad i=2,3,\ldots \tag{4.16}$$

Thus the $\phi_i(t_1;a)$ are arbitrary except that they are to be orthonormal to one another and to $\phi_1(t_1;a)$. One such set is a set generated from $\{\psi_i(t;a)\}$ where

$$\psi_i(t;a) = e^{iat} \qquad i=2,3,\ldots \tag{4.17}$$

By applying the Gram-Schmidt orthonormalization scheme to $\psi_i(t, a)$ we generate the desired complete orthonormal set.† The set $\{\phi_i(t, a)\}$ is now a complete orthonormal set. The characteristic numbers $\mu_i(a)$ are

$$\mu_1(a) = \lambda_1(a) + N_o, \qquad \lambda_1(a) = V_o^2 \frac{e^{2aT}-1}{2a} \tag{4.18}$$

$$\mu_j(a) = N_o, \qquad j=2,3,\ldots$$

B. Expansion of $z(t)$ in terms of the orthonormal (O.N.) set $\{\phi_i\}$.

---

† A complete orthonormal set exists as a result of the integrable square property of $\phi_i(t, a)$ and the positive definiteness of $R_{z|a}(t_1, t_2)$ (Reference 10, p. 374).

Proof to Theorem 4.1 (Continued)

Let

$$z_i = \int_0^T z(t)\, \phi_i(t;a)\, dt \tag{4.19}$$

then (Reference 10, p. 36)

$$z(t) = \underset{n\to\infty}{\mathrm{l.i.m.}} \sum_{i=1}^{n} z_i\, \phi_i(t;a) \tag{4.20}$$

where equality is in the "limit in the mean" sense.

C. Estimate of "a" based on z_i, $i=1, 2, \ldots$.

Let

$$a_N \equiv E[a \mid z_1, \ldots, z_n] = \int a\, p(a \mid z_1, \ldots, z_n)\, da \tag{4.21}$$

or

$$a_N = \frac{\int a\, p(a)\, p(z_1, z_2, \ldots, z_n \mid a)\, da}{\int p(a)\, p(z_1, z_2, \ldots, z_n \mid a)\, da} \tag{4.22}$$

For given a, the set $\{z_i\}$ is a set of uncorrelated zero mean gaussian random variables, i.e., they are a set of independent gaussian random variables.

$$\text{Thus,}\quad p(z_1, z_2, \ldots, z_n \mid a) = \prod_{i=1}^{n} p(z_i \mid a) \tag{4.23}$$

Since $E(z_i \mid a) = 0 \quad \forall\, i$

$$E(z_1^2 \mid a) = \int_0^T\!\!\int_0^T [R_{x|a}(t_1, t_2) + N_o\, \delta(t_1 - t_2)]\, \phi_1(t_1;a)\, \phi_2(t_2;a)\, dt_1\, dt_2$$

$$= \lambda_1(a) + N_o$$

$$E(z_j^2 \mid a) = N_o$$

$$E(z_i z_j \mid a) = 0 \qquad i \neq j \tag{4.24}$$

Proof to Theorem 4.1 (Continued)

then

$$p(z_1, \ldots, z_n | a) = \frac{e^{-\frac{1}{2}\frac{z_1^2}{\lambda_1 + N_o}} \; e^{-\frac{1}{2N_o}\sum_{i=2}^{n} z_i^2}}{(2\pi)^{n/2}\left(\lambda_1(a) + N_o\right)^{\frac{1}{2}} N_o^{\frac{n-1}{2}}} \tag{4.25}$$

Set

$$h_n(a) = \frac{z_1^2}{\lambda_1 + N_o} + \frac{1}{N_o}\sum_{i=2}^{n} z_i^2 \tag{4.26}$$

By adding and subtracting $\frac{z_1^2}{N_o}$, there results

$$h_n(a) = \frac{-\lambda_1}{N_o(N_o+\lambda_1)} + \frac{1}{N_o}\sum_{i=1}^{n} z_i^2 \tag{4.27}$$

Substituting Equations (4.25) and (4.27) into Equation (4.22), and cancelling terms in the numerator and denominator which do not depend upon a, yields

$$\hat{a}_n = \frac{\displaystyle\int a\, p(a) \frac{e^{+\frac{1}{2}\left\{\frac{\lambda_1(a)}{N_o\left(N_o+\lambda_1(a)\right)} z_1^2(a) - \frac{1}{N_o}\sum_{i=1}^{n} z_i^2(a)\right\}}}{(2\pi)^{n/2}\left(N_o+\lambda_1(a)\right)^{\frac{1}{2}} N_o^{\frac{n-1}{2}}}\, da}{\displaystyle\int p(a) \frac{e^{+\frac{1}{2}\left\{\frac{\lambda_1(a)}{N_o(N_o+\lambda)} z_1^2(a) - \frac{1}{N_o}\sum_{i=1} z_i^2(a)\right\}}}{(2\pi)^{n/2}\left(N_o+\lambda_1(a)\right)^{\frac{1}{2}} N_o^{\frac{n-1}{2}}}\, da} \tag{4.28}$$

Proof to Theorem 4.1 (Continued)

$$= \frac{\int a\, p(a)\, e^{+\frac{1}{2}\left\{\frac{\lambda_1 z_1^2(a)}{N_o^2\left(1+\frac{\lambda_1}{N_o}\right)} - \frac{1}{N_o}\sum_{i=1}^{n} z_i^2(a) - \log\left(1+\frac{\lambda_1}{N_o}\right)\right\}} da}{\int p(a)\, e^{+\frac{1}{2}\left\{\frac{\lambda_1 z_1^2(a)}{N_o^2\left(1+\frac{\lambda_1}{N_o}\right)} - \frac{1}{N_o}\sum_{i=1}^{n} z_i^2(a) - \log\left(1+\frac{\lambda_1}{N_o}\right)\right\}} da} \tag{4.29}$$

D. Limiting value of a_n .

The only term that depends on n is the sum of the z_i^2 in the exponential. Taking the limit, yields

$$\underset{n\to\infty}{\mathrm{l.i.m}} \sum_{i=1}^{n} z_i^2(a) = \int_0^T z^2(t)\, dt \, . \tag{4.30}$$

The right-hand side is bounded for

$$E\int_0^T z^2(t)\,dt = E\int_0^T x^2(t)\,dt + N_o T = V_o^2 E\left[\frac{e^{2aT}-1}{2a}\right] + NT < \infty \tag{4.31}$$

and

$$E\left[\int_0^T z^2(t)\, dt\right]^2 < \infty \tag{4.32}$$

since $T < \infty$.

$$\therefore\ \hat{a} = \lim_{n\to\infty} \hat{a}_n$$

$$= \frac{\int a\, p(a)\, e^{+\frac{1}{2}\left\{\frac{\lambda/N_o}{1+\lambda/N_o} z_1^2(a) - \log\left(1+\frac{\lambda}{N_o}\right) - \int_0^T z^2(t)\,dt\right\}} da}{\int p(a)\, e^{+\frac{1}{2}\left\{\frac{\lambda/N_o}{1+\lambda/N_o} z_1^2(a) - \log\left(1+\lambda/N_o\right) - \int_0^T z^2(t)\,dt\right\}} da} \tag{4.33}$$

Proof to Theorem 4.1 (Continued)

$z(t)$ is a realization; it is independent of a since $z(t)$ is given over $0 \le t \le T$. Cancelling $\int_0^T z^2(t)\,dt$ from the exponentials of the numerator and denominator and then substituting for $\lambda_1 z_1(a)$ yields

$$\hat{a}_n = \frac{\displaystyle\int a\,p(a)\,e^{+\frac{1}{2}\left\{\frac{V_o^2/N_o^2}{1+\frac{V_o^2}{N_o}\left(\frac{e^{2aT}-1}{2a}\right)}\left[\int_0^T z(t)\,e^{at}\,dt\right]^2 - \log\left[1+\frac{V_o^2}{N_o}\left(\frac{e^{2aT}-1}{2a}\right)\right]\right\}}\,da}{\displaystyle\int p(a)\,e^{+\frac{1}{2}\left\{\frac{V_o^2/N_o^2}{1+\frac{V_o^2}{N_o}\left(\frac{e^{2aT}-1}{2a}\right)}\left[\int_0^T z(t)\,e^{at}\,dt\right]^2 - \log\left[1+\frac{V_o^2}{N_o}\left(\frac{e^{2aT}-1}{2a}\right)\right]\right\}}\,da} \tag{4.34}$$

which is the Bayes estimate or conditional expectation of a given $z(t)$, $0 \le t \le T$. This proves the main part of the theorem.

E. When $T = 0$, the exponent is zero, and so

$$\hat{a} = \frac{\int a\,p(a)\,da}{\int p(a)\,da} = \bar{a}\,. \tag{4.35}$$

F. When $\frac{V_o^2}{N_o} << 1$, $0 < |aT| << |$ and

$$\frac{V_o^2}{N_o}\left[\int_0^T z(t)\,e^{at}\,dt\right]^2 << 1$$

Proof to Theorem 4.1 (Continued)

then the exponential may be expanded as $e^{\Delta} = 1 + \Delta + \frac{\Delta^2}{2!} + \cdots \approx 1 + \Delta$ where $|\Delta| << 1$. Hence,

$$\frac{e^{2aT}-1}{2a} \approx T \;,$$

$$\log\left\{1 + \frac{V_o^2}{N_o}\left(\frac{e^{2aT}-1}{2a}\right)\right\} \approx \frac{V_o^2 T}{N_o}$$

and so Equation (4.34) becomes

$$\hat{a} \approx \frac{\bar{a} + \frac{V_o^2}{2N_o}\left(\frac{1}{1 + \frac{V_o^2 T}{N_o}}\right)\int_0^T\int_0^T z(t_1)\, z(t_2)\, E\left\{a e^{a(t_1+t_2)}\right\} dt_1\, dt_2}{1 + \frac{V_o^2}{2N_o}\left(\frac{1}{1 + \frac{V_o^2 T}{N_o}}\right)\int_0^T\int_0^T z(t_1)\, z(t_2)\, E\left\{e^{a(t_1+t_2)}\right\} dt_1\, dt_2} \tag{4.36}$$

$$\approx \bar{a} + \int_0^T\int_0^T z(t_1)\, W(t_1,t_2)\, z(t_2)\, dt_1\, dt_2$$

where

$$W(t_1,t_2) = \frac{V_o^2}{2N_o^2}\left(\frac{1}{1 + \frac{V_o^2 T}{N_o}}\right) E\left\{(a-\bar{a})\, e^{a(t_1-t_2)}\right\}$$

This says the low "signal-to-noise" Bayes estimate is a quadratic estimator!

If in addition,

a is gaussian with mean $\bar{a}$ and variance σ

Proof to Theorem 4.1 (Continued)

then

$$W(t_1, t_2) = \frac{V_o^2 \sigma^2}{2N_o^2} \left(\frac{1}{1 + \frac{V_o^2 T}{N_o}} \right) (t_1 + t_2)\, e^{\bar{a}(t_1+t_2)}\, e^{\frac{\sigma^2}{2}(t_1+t_2)^2}$$

and this is a symmetric inseparable kernel.

Q.E.D.

4.3 Theorem 4.2

Given

the same assumptions as given in Theorem 4.1, except that x_o the initial condition on $x(t)$, has a mean not necessarily zero

then

$$\hat{a} = \frac{\int a\, p(a)\, e^{+(1/2)g(a)}\, da}{\int p(a)\, e^{+(1/2)g(a)}\, da}$$

where

$$g(a) = \frac{\frac{V_o^2}{N_o^2}\left[\int_0^T z(t)\, e^{at}\, dt\right]^2 + 2\frac{\bar{x}_o}{N_o}\left[\int_0^T z(t)\, e^{at}\, dt\right] - \frac{\bar{x}_o^2}{N_o}\left[\frac{e^{2aT}-1}{2a}\right]}{1 + \frac{V_o^2}{N_o}} - h(a)$$

and

$$h(a) = \log\left[1 + \frac{V_o^2}{N_o}\left(\frac{e^{2aT}-1}{2a}\right)\right] .$$

(i) If $T = 0$, $\hat{a} = E\, a = \bar{a}$.

(ii) If $0 < |aT| << 1$,

$$|\bar{x}_o| << 1, \ \frac{V_o^2}{N_o} << 1, \ \left|\int z(t)\, e^{at}\, dt\right| < 1$$

then

$$\hat{a} \approx \bar{a} + \int_0^T z(t)\, w_1(t)\, dt + \int_0^T\int_0^T z(t_1)\, w_2(t_1, t_2)\, z(t_2)\, dt_1\, dt_2$$

where

$$w_1(t_1) = \frac{\bar{x}_o}{N_o} E(a-\bar{a})\, e^{at}$$

$$w_2(t_1, t_2) = \frac{V_o^2}{2N_o^2} E(a-\bar{a})\, e^{a(t_1+t_2)} .$$

In other words, for low $\frac{V_o^2}{N_o^2}$ and $\frac{\bar{x}_o}{N_o}$, and short sampling time, the Bayes estimate is a linear plus quadratic weighting of the observables.

Proof to Theorem 4.2

The arguments follow identically as in Theorem 4.1, except

$$E[z_i(a)|a] = \begin{cases} \frac{\bar{x}_o}{k_1} \quad \text{where} \quad \frac{1}{k_1} = \left[\frac{e^{2aT}-1}{2a}\right]^{\frac{1}{2}} & i = 1 \\ 0 & i \neq 1 \end{cases}$$

Let $\bar{z}_1 = \frac{\bar{x}_o}{k_1}$.

Then

Proof to Theorem 4.2 (Continued)

$$p(z_1, \ldots, z_n | a) = \frac{e^{-\frac{1}{2}\left\{\frac{(z_1-\bar{z}_1)^2}{\lambda_1+N_o} + \frac{1}{N_o}\sum_{i=2}^{n} z_i^2\right\}}}{(2\pi)^{n/2}(\lambda_1+N_o)^{\frac{1}{2}} N_o^{\frac{n-1}{2}}}$$

Adding and subtracting $\frac{z_1^2}{N_o}$ in the exponential yields

$$\Gamma_n(a) = \frac{-\frac{\lambda}{N_o} z_1^2 - 2\bar{z}_1 z_1 + (\bar{z}_1)^2}{N_o\left(1+\frac{\lambda}{N_o}\right)} + \frac{1}{N_o}\sum_1^n z_i^2$$

Cancelling the terms $N_o^{\frac{n-1}{2}}$ and $(2\pi)^{n/2}$ in numerator and denominator yields (as in Theorem 4.1)

$$\lim_{n\to\infty} \sum_{i=1} z_i^2 = \int_0^T z^2(t)\,dt$$

and so

$$\Gamma_n(a) \to \Gamma(a)$$

$$\hat{a}_n(a) \to \hat{a}$$

Now $\int_o^T z^2(t)\,dt$ is independent of a, since z(t) is a specific realization and is given. Hence it may be cancelled from the exponentials. Hence,

$$\hat{a} = \frac{\int a\, p(a)\, e^{+\frac{1}{2}g(a)}\, da}{\int p(a)\, e^{+\frac{1}{2}g(a)}\, da},$$

Proof to Theorem 4.2 (Continued)

where

$$g(a) = \frac{\frac{\lambda}{N_o^2} z_1^2 + 2\frac{\bar{z}_1}{N_o} z_1 - \frac{(\bar{z}_1)^2}{N_o}}{1 + \frac{\lambda}{N_o}} - h(a)$$

$$= \frac{\frac{V_o^2}{N_o^2}\left[\int_0^T z(t)\, e^{aT}\, dt\right]^2 + 2\left(\frac{\bar{x}_o}{N_o}\right)\int_0^T z(t)\, e^{at}\, dt - \frac{(\bar{x}_o)^2}{N_o}\left[\frac{e^{2aT}-1}{2a}\right]}{1 + \frac{V_o^2}{N_o}\left(\frac{e^{2aT}-1}{2a}\right)} - h(a)$$

and

$$h(a) = \log\left[1 + \frac{V_o^2}{N_o}\left(\frac{e^{2aT}-1}{2a}\right)\right]$$

Parts (i) and (ii) are obvious.

Q.E.D.

4.4 Theorem 4.3

Given

$$\dot{x}(t) = a\, x(t) + \xi(t) \qquad 0 \le t \le T$$

or

$$x(t) = e^{at}\, x_o + \int_0^t e^{a(t-\rho)}\, \xi(\rho)\, d\rho$$

where

x_o is gaussian $(0, V_o^2)$ and a is a random parameter with density $p(a)$

$$z(t) = x(t) + \eta(t) \qquad 0 \le t \le T$$

If

1. $\xi(t)$ is a white gaussian random process with zero mean and covariance $E\,\xi(\rho_1)\,\xi(\rho_2) = R_o\,\delta(\rho_1-\rho_2)$.

2. $\eta(t)$ is a zero mean white gaussian random process with $E\,\eta(t_1)\,\eta(t_2) = N_o\,\delta(t_1-t_2)$ and uncorrelated with $\xi(t)$.

3. All random terms are independent of each other.

Then

$$a = \frac{\int a\,p(a)\,e^{+\frac{1}{2}Q(a)}\,da}{\int p(a)\,e^{+\frac{1}{2}Q(a)}\,da}$$

where

1. $$Q(a) = \int_0^T\int_0^T z(t_1)\,W\left(t_1,t_2;a,\frac{1}{N_o}\right)z(t_2)\,dt_1\,dt_2 - \frac{1}{N_o}\int_0^T z^2(t)\,dt + \log D\left(a,\frac{1}{N_o}\right)$$

2. $W(t_1, t_2, a, S)$ satisfies the Fredholm Integral Equation[8, 14, 18] of the second kind:

$$W(t_1,t_2;a,S) + S\int_0^T R_{x|a}(t_1,t_3)\,W(t_3,t_2;a,S)\,dt_3 = S\,R_{x|a}(t_1,t_2) \qquad 0 \le t_1,t_2 \le T$$

where $S = \frac{1}{N_o}$

3. $$R_{x|a}(t_1,t_2) = \left(V_o^2 + \frac{R_o}{2a}\right)e^{a(t_1+t_2)} - \frac{R_o}{2a}\,e^{a|t_2-t_1|}$$

4. $D\left(a;\frac{1}{N_o}\right)$ = Fredholm determinant

$$= \exp\left\{\int_0^{1/N_o} \frac{dS}{S}\int_0^T W(t,t;a,S)\,dt\right\}$$

Proof to Theorem 4.3

From the given statistics, it follows

$$R_{x|a}(t_1,t_2) = E\left[x(t_1)x(t_2)|a\right] = V_o^2\, e^{a(t_1+t_2)}$$

$$+\int_0^T\int_0^T e^{a(t_1-p_1)}\, e^{a(t_2-p_2)}\; E\,\xi(p_1)\,\xi(p_2)\,dp_1\,dp_2$$

$$= \left(V_o^2 + \frac{R_o}{2a}\right)e^{a(t_1+t_2)} - \frac{R_o}{2a}\, e^{a|t_2-t_1|}\,.$$

$R_{x|a}(t_1,t_2)$ is symmetric by inspection. If $R_{x|a}(t_1,t_2)$ is positive definite, i.e.,

$$\int_0^T\int_0^T R_{x|a}(t_1,t_2)\,g(t_1)\,g(t_2)\,dt_1\,dt_2 > 0$$

for all integrable square $g(t)$, $t \in [0,T]$ such that $\int_0^T g^2(T)\,dt > 0$,

then the conditions of Davenport and Root, Theorem 8, Page 374 are satisfied and so the orthonormal set of eigenfunctions which satisfy

$$\int_0^T R_{x|a}(t_1,t_2)\,\phi_i(t_2\ a)\,dt_2 = \lambda_i(a)\ \phi_i(t_1,a) \qquad 0 \le t \le T$$

forms a complete orthonormal set. Index i is such that

$$\lambda_1 \ge \lambda_2 \ge \lambda_3 \ge \cdots$$

Proof to Theorem 4.3 (Continued)

If $R_{x|a}(t_1, t_2)$ is positive semi-definite, then the closure of the set must be included to form a closed set $\{\phi_i\}$ which spans the space. $z(t)$ is expressible, then, as

$$z(t) = \underset{n\to\infty}{\text{l.i.m.}} \sum_{i=1}^{n} z_i(a)\ \phi_i(t;a)$$

where

$$z_i = z_i(a) = \int_0^T z(t)\ \phi_i(t, a)\ dt$$

Consider the first n, z_i's. The sequence $\{z_i\}$ is an independent gaussian sequence with mean 0 and variance $\lambda_i + N_o$ for a given "a". Thus

$$p(z_1, z_2, \ldots, z_n | a) = \frac{e^{-\frac{1}{2}\sum_{i=1}^{n}\left[\frac{z_1^2}{\lambda_i + N_o} + \log\left(1 + \frac{\lambda_i}{N_o}\right) + \log N_o\right]}}{(2\pi)^{n/2}}$$

The Bayes estimate based on $z_1, \ldots, z_n$ is

$$\hat{a}_n = \frac{\int a\ p(a)\ e^{-\frac{1}{2} q_n(a)}\ da}{\int p(a)\ e^{-\frac{1}{2} q_n(a)}\ da}$$

where

$$q_n(a) = \sum_{i=1}^{n} \frac{z_i^2}{N_o + \lambda_i(a)} + \log D_n\left(a; \frac{1}{N_o}\right)$$

$$\log D_n\left(a, \frac{1}{N_o}\right) = \sum_{i=1}^{n} \log\left(1 + \frac{\lambda_i(a)}{N}\right)$$

Proof to Theorem 4.3 (Continued)

Substitute for z_1^2, i.e.,

$$q_n(a) =$$

$$\int_0^T\int_0^T z(t_1)\,z(t_2)\left\{\frac{1}{N_o}\sum_{i=1}^{n}\left[1-\frac{\lambda_i}{\lambda_i+N_o}\right]\phi_i(t_1;a)\,\phi_i(t_2;a)\right\}dt_1\,dt_2 +$$

$$+\log D_n\left(a;\frac{1}{N_o}\right)$$

Let

$$g_n(t_1,t_2,a) = \sum_{i=1}^{n}\left[1-\frac{\lambda_i}{\lambda_i+N_o}\right]\phi_i(t_1;a)\;\phi_i(t_2;a) \quad .$$

Let $n\to\infty$

$$g(t_1,t_2;a) = \underset{n\to\infty}{\text{l.i.m.}}\; g_n(t_1,t_2;a)$$

$$= \frac{1}{N_o}\,\delta(t_1-t_2) - W\left(t_1,t_2;a,\frac{1}{N_o}\right)$$

where $W\left(t_1,t_2;a,\frac{1}{N_o}\right)$ has the eigenfunction expansion

$$W\left(t_1,t_2;a,\frac{1}{N_o}\right) = \frac{1}{N_o^2}\sum_{i=1}^{\infty}\frac{\lambda_i}{1+\dfrac{\lambda_i}{N_o}}\;\phi_i(t_1)\;\phi_i(t_2)$$

and satisfies the Fredholm integral equation of the second kind

$$W\left(t_1,t_2;a,\frac{1}{N_o}\right)+\frac{1}{N_o}\int_0^T R_{x|a}(t_1,t_3)\,W\left(t_3,t_2;a,\frac{1}{N_o}\right)dt_3$$

$$= \frac{1}{N_o^2}\,R_{x|a}(t_1,t_2) \;.$$

This is easily seen by substituting the eigenfunction expansion for $W\left(t_1,t_2;a,\frac{1}{N_o}\right)$ into the integral equation and noting

Proof to Theorem 4.3 (Continued)

$$R_{x|a}(t_1, t_2) = \sum_{i=1}^{\infty} \lambda_i \, \phi_i(t_1) \, \phi_i(t_2).$$

Let

$$D\left(a, \frac{1}{N_o}\right) = \lim_{n \to \infty} D_n\left(a, \frac{1}{N_o}\right)$$

where

$D\left(a, \frac{1}{N_o}\right)$ is the Fredholm determinant.[†]

$$\ln D\left(a, \frac{1}{N_o}\right) = \sum_{n=1}^{\infty} \ln\left(1 + \frac{\lambda_i(a)}{N}\right) = \sum_{i=1}^{\infty} \int_0^{1/N_o} \frac{\lambda_i(a)}{1+S\,\lambda_i(a)} \, dS$$

If $t_1 = t_2$ in the equation for $W(t_1, t_2; a, S)$, then

$$\int_0^T W(t, t; a, S)\, dt = S^2 \sum_{i=1}^{\infty} \frac{\lambda_2}{1+\lambda_i} \int_0^T \phi_i^2(t)\, dt = S^2 \sum_{i=1}^{\infty} \frac{\lambda_i}{1+\lambda_2 S}$$

and

$$D\left(a, \frac{1}{N_o}\right) = \exp \int_0^{1/N_o} \frac{dS}{S^2} \int_0^T W(t, t; a, S)\, dt$$

Q.E.D.

4.5 Theorem 4.4

Given

same as Theorem 4.3 except $E\, x_o = \bar{x}_o$ is not necessarily zero, i.e., $x_o \in \text{gaussian}\left(\bar{x}_o, V_o^2\right)$.

Then

[†]See Reference 18, pp. 310-311 or Reference 8, Chapter III.

$$\hat{a} = \frac{\int a\, p(a)\, e^{+\frac{1}{2}Q(a)}\, da}{\int p(a)\, e^{+\frac{1}{2}Q(a)}\, da}$$

where

1. $Q(a) =$

$$\int_0^T\int_0^T \left[z(t_1)-e^{at_1}\bar{x}_o\right] W\left(t_1, t_2; a, \frac{1}{N}\right)\left[z(t_2)-e^{at_2}\bar{x}_o\right] dt_1\, dt_2$$

$$-\int_0^T \left[z(t)-e^{at}\bar{x}_o\right]^2 dt - \log D\left(a, \frac{1}{N_o}\right)$$

2. $W\left(t_1, t_2; a, \frac{1}{N}\right)$ satisfies the Fredholm integral equation of the second kind:

$$W(t_1, t_2; a, s)+S\int_0^T R_{x|a}(t_1, t_3)\, W(t_3, t_2; a, s)\, dt_3$$

$$= S^2\, R_{x|a}(t_1, t_2)$$

where $S = \frac{1}{N_o}$ and $0 \le (t_1, t_2) \le T$.

3. $$R_{x|a}(t_1, t_2) = \left(V_o^2 + \frac{R_o}{2a}\right) e^{a(t_1+t_2)} - \frac{R_o}{2a}\, e^{a|t_2-t_1|}$$

4. $D\left(a; \frac{1}{N_o}\right) =$ Fredholm Determinant

$$= \exp\left\{\int_0^{1/N_o} \frac{dS}{S^2}\int_0^T W(t, t; a, s)\, dt\right\}$$

Proof to Theorem 4.4

Follows directly the steps outlined in Theorem 4.3 except

$$E[z(t)|a] = \bar{x}_o e^{at} = \overline{z(t)}$$

and so $z(t)$ of Theorem 4.3 is replaced by $z(t) - \overline{z(t)}$.

For example

$$\sum_{i=1}^{n} \frac{(z_i - \bar{z}_i)^2}{\lambda_i + N_o} =$$

$$\sum_{i=1}^{n} \int_0^T \int_0^T [z(t_1) - \overline{z(t_1)}][z(t_2) - \overline{z(t_2)}] \frac{\phi_i(t_1\ a)\,\phi_i(t_2\ a)}{\lambda_i + N_o} dt_1\, dt_2$$

and so forth.

4.6 Generalizations and Comments

The general Bayes estimate of the scalar plant a was developed for the case where white disturbing gaussian noise is acting. As far as the author knows, the expressions developed are new. Recently Professor R.E. Mortensen of the University of California, Los Angeles, developed expressions[†] identical to the results of Theorems 4.1 and 4.2, using a completely different approach. He began with the stochastic partial differential equation derived in his Ph.D. dissertation.[33] He converted the stochastic partial differential equations to a standard partial differential equation by adding the "Ito correction factor"[41] but modified for partial differential equations.[21, 22] Fortunately, the partial differential equation that resulted was of first order and quasi-linear. It could be readily solved by standard techniques as in say Hildebrand.[19]

The problem of estimating a is far more complex whenever disturbing noise is acting. In that case the partial differential

† Unpublished correspondence.

equation of Mortensen is quasi-linear and of second order. The solutions given in Theorems 4.3 and 4.4 await verifications by the method of partial differential equations. It should be pointed out that the nonlinear estimation schemes of Stratonovitch,[38] Kushner,[26] and Fisher,[13] result in nonlinear partial differential equations. This is because Mortensen refers the normalization required in order to obtain a probability density until after the solution to the PDE is obtained. In this way he is able to solve these problems at all.

Nevertheless the general case where disturbing noise is present has not been solved by solving the PDE.

It should be observed that expressions obtained in Theorems 4.1 through 4.4 could obviously be extended to the case of estimating a function of a. This is accomplished simply by replacing "a" by f(a) in the a following after the integral sign:

$$\hat{f}(a) = \int f(a)\, p[a \mid z(t), \quad 0 \le t \le T]\ da.$$

The expression derived in Theorems 4.1 through 4.4 are not readily implemented. It appears possible however that a Monte Carlo scheme could be used on a hybrid analog digital computer. One way would be to choose "a" and x_o discretely from an approximate random noise source. Then substitute that value of a into the analog computer to obtain a simulation of the solution of the dynamical system. The solution would be placed in the exponential and weighted appropriately, depending upon the distribution of a and integrated numerically. In this way the desirable features of the analog and the digital computers are used.

Another approach is to approximate the Bayes estimate by a continuous polynomial weighting of the observation. Such polynomial weighting will be discussed next.

CHAPTER V

CONTINUOUS POLYNOMIAL ESTIMATORS

5.1 Description

By the optimum polynomial estimate of degree (say m) is meant a polynomial of the form

$$\hat{a}_m = W^{(o)} + \int_0^T W^{(1)}(t)\, z(t)\, dt + \int_0^T\int_0^T W^{(2)}(t_1, t_2)\, z(t_1)\, z(t_2)\, dt_1\, dt_2 + \cdots +$$

$$+ \underbrace{\int_0^T\int_0^T\int_0^T \cdots \int_0^T}_{m} W^{(n)}(t_1, t_2, \ldots, t_m) \prod_{i=1}^{m} z(t_i)\, dt_i, \tag{5.1}$$

or abstractly the homogenous polynomial

$$\hat{a}_m = \underbrace{\int\int\cdots\int}_{m} K(t_1, \ldots, t_m) \prod_{i=1}^{m} z(t_i)\, dt_i, \tag{5.2}$$

which minimizes the mean square error between it and the Bayes conditional mean $\hat{a}_B = E[a\,|\,z(t), \quad 0 \leq t \leq T]$.

The machinery and notation used in this chapter follows to some extent the recent work of A.V. Balakrishnan.[3,4]

5.2 Sufficient Condition for Optimality

A sufficient condition for polynomial estimators of the form given in Equations (5.1) or (5.2) to minimize the mean square error is that

$$E(a-\hat{a}_m) = 0 \tag{5.3}$$

$$E(a-\hat{a}_m) \prod_{i=1}^{k} z(t_i) = 0 \quad \text{for all} \quad k = 1, 2, \ldots, m$$

$$\text{and} \quad 0 \leq t_i \leq T \tag{5.4}$$

This can be seen by multiplying Equations (5.3) and (5.4) by $W^{(k)}(t_1, t_2, \ldots t_k)$ and integrating $t_1, t_2, \ldots, t_k$ each over $[0, T]$. The sum of these equations is the equation $E(a-\hat{a}_m)\,\hat{a}_m = 0$ and this is the familiar condition for an orthgonal projection of a into the space of the polynomials of degree n. Such a projection minimizes the mean square error $E(a-\hat{a}_m)^2$ among the class of estimators having the form given by Equation (5.1).

5.3 Integral Equations

If Equation (5.3) is substituted into Equation (5.4), then the following set of n integral equations

$$E(a-\bar{a}) \prod_{i=1}^{k} z(t_i) = \sum_{j=1}^{m} \int_0^T \int_0^T \cdots \int_0^T W^{(j)}(\tau_1, \tau_2, \ldots, \tau_j) \left\{ E\left[\prod_{i=1}^{j} z(\tau_i) \prod_{\ell=1}^{k} z(\tau_\ell)\right] - E\left[\prod_{i=1}^{j} z(\tau_i)\right] E\left[\prod_{\ell=1}^{k} z(\tau_2)\right]\right\} d\tau_1 \ldots d\tau_j \qquad (5.5)$$

where $k = 1, 2, \ldots, m$ is to be solved simultaneously for the m weighting functions $W^{(i)}(t_1, \ldots t_i)$, $i = 1, 2, \ldots m$. $W^{(o)}$ is a function of the m weighting functions and the mean of a.

The dynamical system under consideration is assumed observed in a gaussian white noise environment. Consequently in each of the equations $k = 1, 2, \ldots, m$ one of the weighting functions of Equation (5.4) will be outside the integral(s). Hence the set of integral equations, Equations (5.4), will be a set of Fredholm integral equations of the second kind with one or more variables.

By applying the "Contraction Mapping" theorem to the method of successive approximation,[23,40] a sufficient condition on N_o can be obtained such that the ℓ^{th} iteration

$$W_\ell^{(i)}(t_1, \ldots, t_i) = \lambda^{(i)}(N_o)\, f\, W_{\ell-1}^{(i)}(t_1, \ldots, t_i)$$

converges uniformly to $W^{(i)}(t_1, \ldots, t_i)$.

The specific details are described in the next chapter for the case $m = 2$, i.e., the quadratic estimator.

CHAPTER VI

CONTINUOUS QUADRATIC ESTIMATION

In this chapter, the feasibility of using polynomial weighting of second order, $m = 2$, and henceforth called quadratic weighting, is investigated. The simplest non-trivial case is developed, so that the salient points are not lost in an excessive amount of algebra.

6.1 Integral Equations Specifying the Polynomial Weighting Functions

Consider the dynamical system $\frac{dx(t)}{dt} = a\,x(t), \quad 0 \le t \le T$ with no disturbing noise acting and with $x_o = x(o)$ gaussian with zero mean, $\bar{x}_o = 0$, and variance squared, V_o^2. a is a random parameter with known probability density $p(a)$ and mean $\bar{a}$. Consider the system $x(t)$ observed in white gaussian noise $\eta(t)$ and where $\eta(t)$ is characterized by $E\,\eta(t) = 0$ and $E\,\eta(t_1)\,\eta(t_2) = N_o\,\delta(t_1 - t_2)$. Let $z(t)$ be the observation noise of the form $z(t) = x(t) + \eta(t)$, $0 \le t \le T$. If a is the quadratic estimate of a given $z(t)$, $0 \le t \le T$, it has by definition the form

$$\hat{a}_2 = W^{(o)} + \int_0^T W^{(1)}(t)\,z(t)\,dt + \int_0^T\int_0^T W^{(2)}(t_1, t_2)\,z(t_1)\,z(t_2)\,dt_1\,dt_2 \ldots \qquad (6.1)$$

A sufficient condition on $W^{(o)}$, $W^{(1)}(t)$, and $W^{(2)}(t_1, t_2)$, such that the mean square error $\overline{e_2^2} = E(a - \hat{a}_2)$ is minimized, is that

$$E(a - \hat{a}_2) = 0, \quad E(a - \hat{a}_2)\,z(t_3) = 0, \quad E(a - \hat{a}_2)\,z(t_3)\,z(t_4) = 0 \qquad (6.2)$$

for $0 \le t_3, t_4 \le T$. One needs only to multiply these equations by $W^{(o)}$, $W^{(1)}(t_3)$ and $W^{(2)}(t_3, t_4)$ respectively, integrate from 0 to T as necessary and add in order to obtain the orthogonality condition $E(a - \hat{a}_2)\,\hat{a}_2 = 0$. The weighting function of Equation (6.1) which satisfies Equation (6.2) will now be found.

6.2 Theorem 6.1

If the conditions above are satisfied

then

(i) $$W^{(o)} = E\,a - \int_0^T\int_0^T W^{(2)}(t_1,t_2)\,E\,z(t_1)\,z(t_2)\,dt_1\,dt_2 \tag{6.3}$$

(ii) $$W^{(1)}(t) = 0 \tag{6.4}$$

(iii) $W^{(2)}(t_3,t_4)$ satisfies the integral equation

$$W^{(2)}(t_3,t_4) =$$

$$= \frac{\lambda^2}{2} M_{2o}(t_3,t_4) - \frac{\lambda^2}{2}\iint W^{(2)}(t_1,t_2)\,g_2(t_1,t_2;t_3,t_4)\,dt_1\,dt_2$$

$$-\left\{\int_0^T W^{(2)}(t_1,t_4)\,g_1(t_1,t_3)\,dt_1 + \int_0^T W^{(2)}(t_1,t_3)\,g_1(t_1,t_4)\,dt_1\right.$$

$$\left. + \int_0^T W^{(2)}(t_4,t_2)\,g_1(t_2,t_3)\,dt_2 + \int_0^T W^{(2)}(t_3,t_2)\,g_1(t_2,t_4)\,dt_2\right\} \tag{6.5}$$

where

$$\lambda = \frac{1}{N_o}$$

$$m_{2o}(t_1,t_2) = E(a-\bar{a})\,z(t_1)\,z(t_2)$$

$$g_1(t_1,t_2) = V_o^2\,E\,e^{a(t_1+t_2)}$$

$$g_2(t_1,t_2;t_3,t_4) = 3\,V_o^4\,E\,e^{a(t_1+t_2+t_3+t_4)} - V_o^4\,E\,e^{a(t_1+t_2)}\,E\,e^{a(t_3+t_4)}$$

(iv) The quadratic estimator is

$$\hat{a}_2 = E\,a + \int_0^T\int_0^T W^{(2)}(t_1,t_2)\left[z(t_1)\,z(t_2) - E\,z(t_1)\,z(t_2)\right]dt_1\,dt_2 \tag{6.6}$$

(v) The minimum mean square error is

$$\overline{e_2^2} = \sigma^2 - \int_0^T\int_0^T W^{(2)}(t_1, t_2)\, V_o^2\, E(a-\bar{a})\, e^{a(t_1+t_2)}\, dt_1, dt_2 \qquad (6.7)$$

The proof has been referred to the appendix because it is rather detailed, containing several pages of algebra (see Appendix C).

6.3 Some Remarks with Regard to Theorem 6.1

When N_o is sufficiently large or λ is sufficiently small, $W^2(t_1, t_2)$ may be approximated by

$$W^{(2)}(t_1, t_2) \approx \frac{\lambda^2}{2} M_{2o}(t_1, t_2) = \frac{V_o^2}{2N_o} E(a-m)\, z(t_1)\, z(t_2) \qquad (6.8)$$

and so

$$\text{(i)} \quad \hat{a}_2 \approx E\, a + \int_0^T\int_0^T z(t_1)\, z(t_2) \left[\frac{V_o^2}{2N_o^2} E(a-m)\, z(t_1)\, z(t_2)\right] dt_1\, dt_2 \qquad (6.9)$$

but this is the Bayes estimate for $\frac{V_o^2}{N_o^2} << 1$ and short sampling time, i.e., large noise and short sampling time (see Theorem 4.1).

$$\text{(ii)} \quad \overline{e_2^2} \approx \sigma^2 - \int_0^T\int_0^T \frac{V_o^2}{2N_o^2} M_{2o}^2(t_1, t_2)\, dt_1\, dt_2 \qquad (6.10)$$

and if a is gaussian with mean $\bar{a}$ and variance σ,

$$\overline{e_2^2} \approx \sigma^2 - \frac{V_o^2}{2N_o^2} \sigma^4 \int_0^T\int_0^T e^{2\bar{a}(t_1, t_2)}\, e^{\sigma^2(t_1+t_2)^2}\, dt_1\, dt_2 \qquad (6.11)$$

The double integral in the last expression can be evaluated using the tabulated Dawson's integral (Reference 1, p. 319).

(iii) In the next section, Section 6.4, some sufficient conditions are given on N_o or λ to ensure that the above approximations are valid first approximations of $W^{(2)}(t_1, t_2)$.

6.4 Solutions of Fredholm Integral Equations by the Method of Successive Approximations

In this section sufficient conditions are given on the noise covariance N_o to ensure that the sequence $\{W_n^{(2)}(t_1, t_2)\}$ of successive approximations converges uniformly to the unique solution $W^{(2)}(t_1, t_2)$ of the integral equation, Equation (6.5).†

Let R be the space of integrable square, real continuous functions of two variables (t_1, t_2) each defined over the closed interval $[0, T]$ and with a metric

$$p^2(x, y) = \int_0^T\int_0^T [x(t_1, t_2) - y(t_1, t_2)]^2 \, dt_1 \, dt_2 \tag{6.12}$$

for all points $x, y \in R$. R is complete (a closed subset of a complete subset is a complete space).

6.5 Theorem 6.2

Given

conditions above on Equation (6.5) of Theorem 6.1.

Let

$$G_1 = \max_{t_i, t_j \in [0, T]} g_1(t_i, t_j) = \max_{\substack{t_i, t_j \\ t_i, t_j \in [0, T]}} E\, e^{a(t_i + t_j)} \tag{6.13}$$

---

†Another approach to solving the Fredholm integral equation is described by Luchka.[31] By use of his "method of averaging functional corrections" it appears convergence of his successive approximations is more rapid, i.e., involving fewer iterations, but more computation is required per iteration.

$$G_2 = \max_{\substack{t_i \in [0,T] \\ i = 1,2,3}} g_2(t_1, t_2; t_3, t_4)$$

$$= \max_{t_i \in [0,T]} V_o^4 \left\{ 3\, E\, e^{a(t_1+t_2+t_3+t_4)} - E\, e^{a(t_1+t_2)}\, E\, e^{a(t_3+t_4)} \right\} \tag{6.14}$$

If

$$N_o > \frac{G_2 T}{2 G_1} \frac{1}{\sqrt{1 + \frac{G_2}{2G_1^2}} - 1} \tag{6.15}$$

Then

(i) the sequence $\{W_n^{(2)}(t_3, t_4)\}$ generated by

$$W_o^{(2)}(t_3, t_4) = \frac{1}{2N_o} M_{2o}(t_3, t_4) \tag{6.16}$$

$$W_{n+1}^{(2)}(t_3,t_4) = \frac{1}{2N_o^2} M_{2o}(t_3, t_4)$$

$$- \frac{1}{2N_o} \left\{ \int_0^T W_n^{(2)}(t_1, t_4)\, g_1(t_1, t_3)\, dt_1 + \int_0^T W_n^{(2)}(t_1, t_3)\, g(t_1, t_4)\, dt_1 + \right.$$

$$\left. + \int_0^T W_n^{(2)}(t_4, t_2)\, g_1(t_2, t_3)\, dt_2 + \int_0^T W_n^{(2)}(t_3, t_2)\, g_1(t_2, t_4)\, dt_2 \right\} +$$

$$- \frac{1}{2N_o^2} \int_0^T \int_0^T W_n^{(2)}(t_1, t_2)\, g_2(t_1, t_2; t_3, t_4)\, dt_1\, dt_2 \tag{6.17}$$

converges uniformly to the unique solution $W^{(2)}(t_3, t_4)$ of Equation (6.5).

(ii) If in particular, $\frac{G_2}{2G_1^2} << 1$, then convergence occurs for

$$N_o > 2\, G_1 T$$

Remark

If the first approximation of $W^{(2)}(t_3, t_4)$, given by Equation (6.16) is substituted into Equation (6.6), the optimum quadratic estimate is the high noise $\left(\frac{V_o^2}{N_o} << 1\right)$, short time of observation Bayes estimate. In a sense this is analogous to the "low signal-to-noise" case except that here, the mean $\bar{x}$ is zero and the variance of x_o is small compared to the noise. In short as V_o decreases, the spread of the distribution of x_o about $\bar{x}_o = 0$ decreases, so that any realizations will tend to be near $x_o = 0$. For this reason $\frac{V_o^2}{N_o}$ may be considered a power "signal-to-noise" ratio.

For the case $\bar{x}_o \neq 0$, Equation (6.2) results in two coupled Fredholm equations of the second kind (see Equations C.2 and C.3 of Appendix C). Sufficient conditions can be given on N_o, just as in this proof, so that the sequence of weighting functions $W_k^{(1)}(t)$ and $W_k^{(2)}(t_1, t_2)$ wil converge uniformly to the solution of the coupled equations.

Proof to Theorem 6.2

Let $\lambda = \frac{1}{N_o}$. Set $W_o^{(2)} = W_o^{(2)}(t_1, t_2) = \lambda M_{2o}(t_1, t_2)$ as the initial iteration. Let $\Delta W_{n+1}^{(2)} = W_{n+1}^{(2)} - W_n^{(2)}$. Substitute $W_{n+1}^{(2)}$ and then $W_n^{(n)}$ into Equation (6.17), subtract equations, apply Equation (6.12). Then square, integrate twice over $[0, T]$ and apply Schwartz inequality. The result is

$$\rho_{n+1}^2 \equiv \rho^2(W_{n+1}, W_n) \leq \left(\frac{\lambda T}{2}\right)^2 \left[\lambda G_2 T + 4 G_1\right]^2 \rho_n^2 \ .$$

If $\lambda > 0$ is chosen such that $|\alpha| \equiv \frac{\lambda T}{2}\left[\lambda G_2 T + 4 G_1\right] < 1$,

$$\rho_{n+1} < |\alpha| \, \rho_n \ .$$

Proof to Theorem 6.2 (Continued)

The conditions for the contraction mapping theorem to hold are satisfied. Hence $\rho \to 0$ as $n \to \infty$ and $W_n^{(2)} \to W^{(2)}$ uniformly as $n \to \infty$. Equation (3.39) implies

$$\lambda^2 G_2 T^2 + 4 G_1 T - 2 < 0$$

or since $\lambda = \frac{1}{N_o} > 0$, the positive root is desired.

$$\therefore N_o > \frac{\frac{G_2 T}{2G_1}}{\sqrt{1 + \frac{G_2}{2G_1}} - 1} .$$

If $\frac{G_2}{2G_1} << 1$, the sufficient condition becomes $N_o > GT$.

Q.E.D.

Remarks

1. Convergence is assured even when the noise is assumed sufficiently large.
2. For "a" gaussian with the mean of a negative G_2 and G_1 will have an extreme maximum either at $t_i = 0$ or at $t_i = T$, depending upon the magnitude of $\bar{a}$, σ, and T.
3. Lower bounds on N_o can be developed by considering say the operator G, defined by

$$G\,h = \int_0^T \int_0^T g(t_1, t_2; t_3, t_4)\, h(t_1, t_2)\, dt_1\, dt_2$$

and mapping say $h \in L^p(0, T)$ into $L^p(0, T)$ i.e., let $h_n = G\, h_{n-1}$, $\|G\|$ = norm of G, and

$$\|h\|^2 = \int\int h^2(t_1, t_2)\, dt_1\, dt_2$$

since

$$\|h_{n+1} - h_n\| = \|G^p(h_{n-p} - h_{n-p-1})\|$$
$$\leqq \|G^p\|\ \|h_{n-p} - h_{n-p-1}\|$$

Then

$$\alpha \equiv \sqrt{\|G^p\|} < 1$$

and this gives a tighter bound on N_o.

6.6 Cramer-Rao Inequality[†]

The Cramer-Rao inequality gives the lower bound on the conditional mean, i.e.,

$$E\left[(a-a_2)^2 \middle| a\right] \geq \frac{\left(1 + \frac{\partial b}{\partial a}\right)^2}{I(a)}$$

where

$$I_n(a) = E\left\{\left[\frac{\partial}{\partial a} \log p(z_1, z_2, \ldots, z_n | a)\right]^2 \middle| a\right\}$$

= Information in the sample [Reference 33, p. 218]

$$I(a) = \underset{n\to\infty}{\text{l.i.m.}}\, I_n(a)$$

$z_i = z_i(a)$ are the coefficients of $\int_0^T z(t)\, \phi_i(t;a)\, dt$

and $\{\phi_i(t;a)\}$ are the orthonormal set of eigenfunctions described in Theorem 4.1. For the conditions of Theorem 4.1, the bias is

$$b(a) = E(\hat{a}_2 | a) - a$$

†See References 9, 28, 29, 30.

Hence

$$\left(1 + \frac{\partial b}{\partial a}\right)^2 = V_o^4 \left\{ \int_0^T \int_0^T W_2(\tau_1, \tau_2) \left[(\tau_1 + \tau_2)\, e^{a(\tau_1 + \tau_2)} \right] d\tau_1\, d\tau_2 \right\}^2$$

and so the quadratic estimate is a biased estimate.

In order to determine the lower bound on the conditional mean, it is necessary to know I(a). The term I(a) however causes difficulty. This is because the term $Q_n(a)$ given by

$$Z_n(a) = \log p(z_1, z_2, \ldots, z_n | a)$$

$$= \frac{1}{2} \frac{\lambda_1(a)\, z_1^2(a)}{N_o\left(\lambda_1(a) + N_o\right)} - \frac{1}{2} \log\left(\lambda_1(a) + N_o\right) - \frac{1}{2N_o} \sum_{i=1}^{n} z_i^2 + c_n$$

where

$$\lambda(a) = V_o^2 \frac{e^{2aT} - 1}{2a}$$

$$c_n = -\frac{n}{2} \log 2\pi - \frac{n-1}{2} \log N_o$$

must be differentiated with respect to a, averaged over $z_1, \ldots, z_n$ given a, and then the limit as $n \to \infty$ taken.

Consequently the Cramer-Rao bound is not directly obtainable.

6.7 Synopsis of Chapter VI

In this chapter the optimum minimum variance polynomial estimator of order two was found, i.e., in the sense that it satisfy a Fredholm integral equation of the second kind in two variables. By specifying sufficient conditions on the noise covariance, the method of successive approximations yields a convergent sequence which converges to the optimum quadratic weighting function. The estimator that results reduces to the Bayes estimate when the covariance N_o of the measurement noise is sufficiently large and the time of observation is sufficiently small.

A method for determining the lower bound on the conditional error squared, known as the Cramer-Rao inequality, was sketched. The minimum variance mean square error for the quadratic estimator coincided with the Bayes estimate whenever N_o is large $(N_o >> V_o^2)$ and T is small. The quadratic estimator was shown to be a biased minimum mean-square-error estimator.

CHAPTER VII

MATRIX PLANT

7.1 Introduction

Almost all of the notions developed in the earlier chapters carry over to the vector dynamical system. The only added feature is that the elements in the plant matrix that are to be estimated should be rearranged, so that they form a vector. The same is true of the observables; they should be rearranged into an array. If this is done, then n^{th} order polynomial weighting for both continuous time parameterization and discrete time parameterization carry over. In fact, if the mean of the initial condition on the state $x(t)$ is zero, then linear weighting on the observables does not reduce the mean square error. It is not until higher order weighting occurs that the estimate of the plant parameter(s) improve.

The prior statistics required to specify a matrix of random parameters increases rapidly with the size of the matrix.[16] For example, a 2x2 matrix of 4 elements requires 4 means and 10 covariances; a 3x3 matrix, 9 means and 45 covariances; a 4x4 matrix, 16 means and 136 covariances; and a 5x5 matrix, 25 means and 325 covariances. Most linear physical systems however will not require anywhere near this number because only the last row of elements of the A matrix are random. Under these circumstances the number of statistics required to specify a 2x2, 3x3, 4x4, or 5x5 matrix of gaussian parameters, is respectively, 5, 9, 14, 20. Thus, an application of physical intuition permits an appreciable reduction in the number of quantities required to specify the distribution associated with the A matrix.

7.2 Example of Continuous Quadratic Weighting

To demonstrate the procedure once again, the following single non-trivial vector dynamic system is considered. Assume without loss of generality that no disturbing noise is acting on the dynamical system. (If there were, a few more obvious terms would need to be added to the statistical terms that are already indicated.)

The dynamical system is

$$\ddot{x}(t) + a_1 \dot{x}(t) + a_o x(t) = 0, \quad x(o) = x_o \tag{7.1}$$

It can be rearranged into the vector form

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -a_o & -a_1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

that is

$$\dot{\underline{X}}(t) = A\, \underline{X}(t), \qquad 0 \le t \le T \tag{7.2}$$

a_o and a_1 are the random parameters to be estimated. The measurement system measures $\underline{X}(t)$ but in white gaussian noise environment:

$$\underline{Z}(t) = \underline{X}(t) + \eta(t) = \begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} \tag{7.3}$$

The solution of Equation (7.2) is

$$\underline{X}(t) = e^{At}\, \underline{X}_o \tag{7.4}$$

but

$$e^{At} = \alpha(\lambda_1, \lambda_2, t)\, I + \beta(\lambda_1, \lambda_2, t)\, A\, t \tag{7.5}$$

where λ_1 and λ_2 are roots of $|\lambda I - A| = 0$ and α, β satisfy

$$e^{\lambda_1 t} = \alpha + \beta\, \lambda_1 t$$

$$e^{\lambda_2 t} = \alpha + \beta\, \lambda_2 t$$

The two equations above are solved for α and β and then α and β substituted into Equation (7.5). λ_1, λ_2 are functions of the plant elements a_o, a_1. Consequently, α and β are also. Z(t) then has the form

$$Z(t) = \alpha(a_o, a_1, t)\, I + \beta\,(a_o, a_1, t)\, A\, t + \underline{\eta}(t) \tag{7.6}$$

Next form the vector $\mathfrak{z}^{(2)}(t_1, t_2)$. It is defined by the terms of second degree generated from $\underline{Z}(t_1)$ and $\underline{Z}(t_2)$. Specifically

$$\mathfrak{z}^{(2)}(t_1, t_2) = \begin{bmatrix} z_1(t_1)\, z_1(t_2) \\ z_1(t_1)\, z_2(t_2) \\ z_2(t_1)\, z_1(t_2) \\ z_2(t_1)\, z_2(t_2) \end{bmatrix}_{4x1} \tag{7.7}$$

Let $\underline{b} = \begin{bmatrix} a_o \\ a_1 \end{bmatrix}$. The optimum estimate of $\underline{b}$ (and hence of a_o and a_1) of the form

$$\hat{b} = W^{(o)} + \int_0^T W^{(1)}(t)\, Z(t)\, dt + \int_0^T\!\!\int_0^T W^{(2)}(t_1, t_2)\, \mathfrak{z}^{(2)}(t_1, t_2)\, dt_1\, dt_2 \tag{7.8}$$

is to be found. Z(t) is a 2x1 matrix while $\mathfrak{z}^{(2)}(t_1, t_2)$ is a 4x1 matrix; $W^{(o)}$ is 2x1, $W^{(1)}(t)$ is 2x2, and $W^{(2)}(t_1, t_2)$ is 2x4. Had we not known that two of the elements of A were 0 and 1, then all four elements of A would need to be estimated. This would result in $\underline{b}$ being 4x1; $W^{(o)}$ being 4x1; $W^{(1)}(t)$ being 4x2; and $W^{(2)}(t_1, t_2)$ being 4x4. The result is twice as many parameters would need to be determined. Fortunately this generally will not be the case, for additional information is often available. Returning now to the original problem of Equation (7.1), we find a sufficient condition for an estimator of the form of Equation (7.8) to minimize the mean square error is that

$$E(b-\hat{b}) = 0$$

$$E(b-\hat{b})\, Z^*(\tau) = 0$$

$$E(b-\hat{b})\, \mathfrak{z}^{(2)*}(\tau_1, \tau_2) = 0$$

These are the conditions stated many times in the past. From them one obtains the equations:

$$E(b-\hat{b})\left(Z(\tau)-\overline{Z(\tau)}\right)^* = \int_0^T W^{(1)}(t_1)\left[E\, Z(t_1)\, Z^*(\tau) - E\, Z(t_1)\, E\, Z^*(\tau)\right] dt_1$$

$$+\int_0^T\int_0^T W^{(2)}(t_1, t_2)\left[E\, \mathfrak{z}^{(2)}(t_1, t_2)\, Z^*(\tau) - E\, \mathfrak{z}^{(2)}(t_1, t_2) E Z^*(\tau)\right] dt_1\, dt_2$$

and

$$E(b-\hat{b})\left(\mathfrak{z}^{(2)}(\tau_1, \tau_2) - \overline{\mathfrak{z}^{(2)}(\tau_1, \tau_2)}\right)^* =$$

$$= \int_0^T W^{(1)}(t_1)\left[E\, Z(t_1)\, \mathfrak{z}^{(2)*}(\tau_1, \tau_2) - E\, Z(t_1)\, E\, \mathfrak{z}^{(2)*}(\tau_1, \tau_2)\right] dt_1 +$$

$$+\int_0^T\int_0^T W^{(2)}(t_1, t_2)\Big\{E\, \mathfrak{z}^{(2)}(t_1, t_2)\, \mathfrak{z}^{(2)*}(\tau_1, \tau_2) -$$

$$- E\, \mathfrak{z}^{(2)}(t_1, t_2)\, \mathfrak{z}^{(2)*}(\tau_1, \tau_2)\Big\} dt_1\, dt_2$$

Since the measurement noise is assumed white gaussian, a pair of vector Fredholm equations of the second kind occur. As in Section 6.4 of Chapter VI, sufficient conditions may be found so that the method of successive approximations could be used to find the optimum weighting functions and the minimum mean square error. The details are a straightforward extension of the discussion in Chapter VI. Incidently, the discrete quadratic estimator may be developed, using a method similar to the method used to solve the continuous problem. The extension to higher degree weighting is simply a straightforward application of the techniques already developed.

SUMMARY, RESULTS, AND CONCLUSIONS

Summary

The identification of linear time invariance stochastic dynamical systems is made for the class of systems where the plant is characterized by either a scalar random parameter or by a matrix of random parameters having known distribution. The initial state is assumed gaussianly distributed with known mean and variance. White gaussian disturbing noise with zero mean and known variance is acting on the dynamical system.

The state of the system is observed under the influence of additive white gaussian noise, having zero mean and known covariance. The white gaussian measurement noise is statistically independent of the white gaussian disturbing noise. The problem is to identify the plant parameters such that the mean square error between the plant parameter(s) and the estimate are minimized. The optimum estimate is the conditional mean or Bayes estimate.

Both discrete and continuous systems are investigated. Recursive polynomial estimators are described and some of their asymptotic properties discussed. Both discrete and continuous recursive quadratic estimators are designed and extensions noted.

Results

The Bayes estimate of the scalar plant parameter for the discrete case requires integrating a function which has the ratio of two polynomials in a in the exponent. The order of the polynomials grows as $2(k+1)$ where $k+1$ is the number of observations. The integral is not readily integrable. However, it is shown that the function mentioned above, as well as the Bayes estimate which contains this function, may be approximated as an infinite sum of polynomials in the $k+1$ variables ranging in order from zero to infinity.

If the mean initial state is zero, linear weighting on the observables is shown to give no improvement to the estimate of the plant parameter beyond the initial estimate, i.e., its mean. In short, not until nonlinear weighting on the observables occurs does the estimate improve. IBM 7094 FORTRAN Programs of the linear weighting case and quadratic weighting case indicate the superiority of quadratic weighting. The programs are described in Appendix B. By using recursive estimators, the dimension of the space required reduces from $\frac{(k+1)(k+4)}{2}$ to $k+1$. This means the number of elements of a matrix to be inverted is reduced by $\left[\frac{(k+1)(k+4)}{2}\right]^2$ to $(k+1)^2$. Extensions to higher dimensions are indicated. Figures 2 and 3 show the rate of increase of the required state space as the degree of the polynomial and the number of observation changes.

The Bayes estimate for the scalar random parameter continuous time case is derived (see Theorems 4.1 through 4.4). The integrals to be evaluated appear too difficult to be solved directly, however they could be evaluated by numerical techniques, once the specific realization is available.

Continuous polynomial estimators are developed. Sufficient conditions are given on the weighting functions, so that the mean square error between the plant parameter a and the polynomial estimate a_m is minimized. The conditions result in a set of coupled Fredholm equations of the second kind. For the case where quadratic weighting is used and the mean initial state is zero, sufficient conditions are given on the measurement noise so that the above Fredholm equations may be solved by the method of successive approximations.

The techniques specified above apply directly to linear vector dynamical systems. A second order system with damping is investigated.

Conclusions

Polynomial estimators are specifically useful where noise levels are high and the observation time (or number of samples) is small. They also have the advantage of having their weighting functions predetermined.

REFERENCES

1. Abramowitz, Milton and Irene A. Stegun, "Table of Dawson's Integral," Handbook of Mathematical Functions, National Bureau of Standards, U.S. Government Printing Office, Washington, D.C., 1961, p. 319.

2. Aitchison, J. and J.A.C. Brown, The Lognormal Distribution, Cambridge University Press, Cambridge, England, 1957.

3. Balakrishnan, A.V., "A General Theory of Nonlinear Estimation Problems in Control Systems," J.Math. Analysis and Applications, 8, 4-30, February 1964.

4. Balakrishnan, A.V., "On a Class of Nonlinear Estimation Problems," IEEE Trans. on Information Theory, IT-10(4), 314-320, October 1964.

5. Bass, R.W., V.D. Norum and L. Schwartz, "Optimal Multichannel Nonlinear Filtering," Hughes Aircraft Co., Space Systems Div., Los Angeles, Report No. SSD-50064 R, August 25, 1965.

6. Bucy, R.S., "Nonlinear Filtering Theory," IEEE Trans. on Automatic Control, AC-10(2), 198, April 1965.

7. Cameron, R.H. and W.T. Martin, "The Orthogonal Development of Non-Linear Functionals in Series of Fourier-Hermite Functions," Annals of Mathematics, 48(2), 385-392, April 1947.

8. Courant, R. and D. Hilbert, Methods of Mathematical Physics, Vol. I, First English edition, Interscience Publishers, New York, 1953.

9. Cramer, Harold, Mathematical Methods of Statistics, Princeton University Press, Princeton, N.J., 1946.

10. Davenport, Wilbur B., Jr., and William L. Root, An Introduction to the Theory of Random Signals and Noise, McGraw-Hill, New York, 1958.

11. Doob, J.L., Stochastic Processes, Wiley, New York, 1953.

12. Ferguson, Thomas S., Statistical Inference, Class notes for Statistics 231AB, Department of Mathematics, University of California, Los Angeles, California, 1965-1966.

13. Fisher, James Richard, Optimal Nonlinear Filtering, Ph.D. in Engineering, University of California, Los Angeles, January 1966.

14. Goursat, Edouard, A Course in Mathematical Analysis: Vol. III, Part 2: Integral Equations, Calculus of Variations, (Translated by Howard G. Bergmann), Dover Publications, Inc., New York, 1964.

15. Gray, A.H., Jr. and T.K. Gaughey, "A Controversy in Problems Involving Random Parametric Excitation," J. Math. and Physics, 44(3), 288-296, September 1965.

16. Grenander, Ulf., "Densities for Stochastic Processes," Arkiv for Matematik, 1, 195-277, 1950.

17. Gunckel, T.L., II. "Nonlinear Sequential Estimation," Paper delivered in the two week University Extension Summer class, Modern Random Processes with Control Systems Applications (class directed by Prof. A.R. Stubberud and Prof. E.B. Stear), Department of Engineering, University of California, Los Angeles, July 6-16, 1965.

18. Helstrom, Carl W., Statistical Theory of Signal Detection, Pergamon Press, London, 1960.

19. Hildebrand, F.B., Advanced Calculus for Engineers, Prentice-Hall, New York, 1949.

20. Ho, Yu-Chi and R.C.K. Lee, "A Bayesian Approach to Problems in Stochastic Estimation and Control," Proceedings of the 1965 Joint Automatic Control Conference, Rensselaer Polytechnic Institute, Troy, New York, June 1965, p. 382-387.

21. Ito, Kiyosi, Lectures on Stochastic Processes, Tata Institute of Fundamental Research, Bombay, India, 1961.

22. Ito, K., "On Stochastic Differential Equations," Memoirs of the American Math. Soc., 4, 1-51, 1951.

23. Kolmogorov, A. and S.V. Fomin, Elements of the Theory of Functions and Functional Analysis: Vol. I - Metric and Normed Spaces, 1957, Vol. II - Measure, the Lebesque Integral, Hilbert Space, 1961, Graylock Press, Albany, N.Y.

24. Kopp, Richard E. and Richard J. Orford, "Linear Regression Applied to System Identification for Adaptive Control Systems," AIAA Journal, 1(10), 2300-2306, October 1963.

25. Kumar, K.S.P. and R. Sridhar, "On the Identification of Control Systems by the Quasi-Linearization Method," IEEE Trans. on Auto. Control, AC-9(2), 151-159, April 1964.

26. Kushner, H.J., "On the Differential Equations Satisfied by Conditional Probability Densities of Markov Processes with Applications," SIAM J. of Control, Ser. A, 2(1), 106-119, 1964.

27. Lee, Robert C.K., Optimum Estimation, Identification, and Control, The M.I.T. Press, Cambridge, Mass. 1964.

28. Lehmann, E.L., Notes on the Theory of Estimation, Chapters 1 to 5 (Lectures by E.L. Lehmann 1949-50, notes recorded by Colin Blyth, reprinted Spring 1962). Originally printed September 1960, University of California, Berkeley.

29. Lindgren, B.W., Statistical Theory, Macmillan, New York, 1962.

30. Loeve, Michel, Probability Theory, D. Van Nostrand, Princeton, N.J., 1955.

31. Luchka, A. Yu, The Method of Averaging Functional Corrections: Theory and Applications (Academic Paperbacks), Academic Press, New York, 1965.

32. Masani, P. and N. Wiener, "Nonlinear Prediction," Probability and Statistics: The Harold Cramer Volume (Ed. by Ulf Grenander), Wiley, New York, 1959, p.190-212.

33. Mortensen, Richard Edgar, Optimal Control of Continuous Time Stochastic Systems, Ph.D. in Engineering, University of California, Berkeley, January 1966.

34. Ohap, Richard Frank, A Technique for the Optimal Estimation of the State of a Nonlinear System, Exemplified by an Orbit Determination Problem, M.S. in Engineer, University of California, Los Angeles, June 1964.

35. Polya, G. and G. Szego, Aufgaben und Lehrsatze aus der Analysis, Springer-Verlag, Berlin, 1964.

36. Riesz, Frigyes and Bela Sz.-Nagy, Functional Analysis (translated from the 2nd French edition by Leo. F. Boron), Ungar Publishing Co., New York, 1955.

37. Seal, Charles Elwood, On Control Systems with Measurement Errors of the State Variables, M.S. in Engineering, University of California, Los Angeles, 1964.

38. Stratonovitch, R.L., "Conditional Markov Processes," Theory of Probability and its Applications, 5(2), 156-178, 1962.

39. Stubberud, Allen R., "Optimal Orbit Transfer," Aerospace Corporation, El Segundo Technial Operations, Report No. TOR-269(4130-13)-1, El Segundo, California, June 26, 1964.

40. Whalen, B., "The Contraction Theorem and its Extensions and Applications," Electronics Research Laboratory, Report No. 64-42, University of California, Berkeley, November 11, 1964.

41. Wong, E. and M. Zakai, "On the Relation Between Ordinary and Stochastic Differential Equations," Electronics Research Laboratory, Report No. 63-26, University of California, Berkeley, August 11, 1964.

42. Wonham, W.M., "Stochastic Problems in Optimal Control," IEEE International Convention Record 1963, Part 2, 114-124, 1963.

APPENDIX A

RECURSION EQUATIONS FOR LINEAR ESTIMATOR

The dynamical system is

$$x(k+1) = a\,x(k) + \xi(k) = a^k x_o + \sum_{j=1}^{k} a^{k-j}\,\xi_{j-1} \tag{A.1}$$

$$z(k) = x(k) + \eta(k) \tag{A.2}$$

Let

$$y(k) = z(k) - E\,z(k) \tag{A.3}$$

and let

$$\Delta y(k) = y(k) - \hat{E}\Big(y(k)\,\Big|\,y(o), y(1), \ldots, y(k-1)\Big)$$

where the circumflex over E denotes the optimum <u>linear</u> estimate of $y(k)$ when $y(o), y(1), \ldots, y(k-1)$ is known (see Doob, Reference 11). Then

$$\Delta y(k) = y(k) - \hat{E}\Big(y(k)\,\Big|\,\Delta y(o), \Delta y(1), \ldots, \Delta y(k-1)\Big) \tag{A.4}$$

Let

$$\Delta Y_1(k-1) = \begin{bmatrix} \Delta y(o) \\ \Delta y(1) \\ \vdots \\ \Delta y(k-1) \end{bmatrix} \tag{A.5}$$

Then $E\,\Delta Y_1(j)\,\Delta Y_1(j)^*$ is a diagonal matrix with diagonal elements $E\,\Delta y(i)\,\Delta y(i)$ and

$$\Delta y(k) = y(k) - \sum_{j=1}^{k}\Big(E\,y(k)\,\Delta y(k-j)\Big)\Big(E\,\Delta y(k-j)\,\Delta y(k-j)\Big)^{-1}\Delta y(k-j) \tag{A.6}$$

Note $E\,y(i)\,\Delta y(j) = 0$ for all $j \geq i$. Let

$$m_1(k-j, k) \equiv E\,\Delta y(k-j)\,y(k) \tag{A.7}$$

$$p_1(j) \equiv E\,\Delta y(j)\,\Delta y(j) = m_1(j, j) \tag{A.8}$$

$$c_1(k) \equiv E\,a\,\Delta y(k) \tag{A.9}$$

Substituting Equation (A.6) into Equations (A.7) and (A.9) gives

$$m_1(k-j,k) = E\,y(k-j)\,y(k)$$

$$- \sum_{\ell=0}^{k-j-1} m_1^*(k-j,k-j-1-\ell)\, p_1^{-1}(k-j-1-\ell)\, m(k-j-1-\ell,k) \tag{A.10}$$

$$c_1(k) = E\,a\,y(k) - \sum_{\ell=1}^{k} c_1^*(k-\ell)\, p_1^{-1}(k-\ell)\, m(k-\ell,k) \tag{A.11}$$

The optimum linear estimate of a given $z_o \ldots z_k$, denoted by $\hat{a}(k)$, is

$$\hat{a}_1(k) = \hat{E}[a\,|\,z(o), z(1), \ldots, z(k)] = \hat{E}[a\,|\,\Delta y(o), \ldots, \Delta y(k)]$$

That is,

$$\hat{a}_1(k) = \bar{a} + \left(\left(E\,a\,\Delta Y(k)\right)\right) \left(E\,\Delta Y(k)\,\Delta Y^*(k)\right)^{-1} \Delta Y(k)$$

$$= \bar{a} + \sum_{j=0}^{k} c_1(k-j)\, m_1^{-1}(k-j,k-j)\, \Delta y(k-j)$$

$$= \hat{a}_1(k-1) + c_1(k)\, m_1^{-1}(k,k)\, \Delta y(k)$$

Similarly, the minimum mean square error resulting from linear weightings is

$$\overline{e_1^2(k)} = E(a-\bar{a})^2 - \left(E\,a\,\Delta y(k)\right)^* \left[E\,\Delta y(k)\,\Delta y(k)\right]^{-1} \left(E\,a\,\Delta y(k)\right)$$

$$= \overline{e_1^2(k-1)} - E\,a\,\Delta y(k)^* \left[E\,\Delta y(k)\,\Delta y(k)\right]^{-1} \left(E\,a\,\Delta y(k)\right)$$

$$= e_1^2(k-1) - c_1(k)\, p_1^{-1}(k)\, c_1(k)$$

and gives the desired relations. An example has been programmed for the IBM 7094 using FORTRAN IV language. The plant parameter a is assumed to be gaussianly distributed. The mean squared error normalized by the variance of a has been plotted. See Figure 1.

APPENDIX B

COMPUTER PROGRAMS

The equations developed in Appendix A were set into FORTRAN IV computer language. Double precision was used where possible. The program generated the required statistical moments and evaluated the necessary terms via recursive relations.

The distribution of the plant parameter was chosen rather arbitrarily to be gaussian with some mean and variance but it could just as easily have been some other distribution with the moments determined either in closed form or by a Monte Carlo scheme.

I. LINEAR ESTIMATOR

The k^{th} moment for a gaussian with mean $\bar{a}$ and variance σ is given by

$$\overline{a^k} = \begin{cases} 1 & k = 0 \\ \bar{a} & k = 1 \\ (\bar{a})^2 + \sigma^2 & k = 2 \\ (\bar{a})^k + \sum_{\ell=1}^{\left[\frac{k}{2}\right]} \left\{ \frac{(k)(k-1)(k-2)\cdots[k-(2\ell-1)]}{\ell!} \left(\frac{\sigma^2}{2}\right)^{\ell} (\bar{a})^{k-2\ell} \right\} & k \geq 3 \end{cases}$$

where the bar over a denotes the expectation and the bracket term above the summation sign, i.e., $\left[\frac{k}{2}\right]$, denotes the largest integer $\leq \frac{k}{2}$. A double precision function was defined which generated $\overline{a^k}$. This subprogram was then used to generate the subprograms E y(i) y(j) and E a y(i); that is

$$E\, y(i)\, y(j) = \begin{cases} \overline{a^3}\,\overline{x_o^2} - \overline{a^2}\,\bar{a}(\bar{x}_o)^2 + R_o \sum_{\ell=1}^{i} \overline{a^{i+j-2\ell}} + N_o\,\delta_{ij} & i \leq j \\ \overline{a^3}\,\overline{x_o^2} - \overline{a^2}\,\bar{a}(\bar{x}_o)^2 + R_o \sum_{\ell=1}^{j} \overline{a^{i+j-2\ell}} + N_o\,\delta_{ij} & j \leq i \end{cases}$$

and

$$E\, a\, y(i) = \begin{cases} 0 & i = 0 \\ (\overline{a^{i+1}} - \bar{a}\,\overline{a^i})\,\overline{x_o} & i \geq 1 \end{cases}$$

The above subprograms were then used in the recursive equation

$$m_1(k-j, k) = E\; y(k-j)\, y(k) - \sum_{\ell=1}^{k-j} m_1(k-j, k-j-\ell)\, p_1^{-1}(k-j-\ell)\, m_1(k-j-\ell, k)$$

with

$$m_1(o, o) \equiv E\; z^2(o) - \big(E\, z(o)\big)^2 = V_o^2 + N_o$$

and

$$p_1(j) = m_1(j, j)\,,$$

and in the recursive equation

$$c(k) = E\, a\, y(k) - \sum_{\ell=1}^{k} c^*(k-\ell)\, p_1^{-1}(k-\ell)\, m(k-\ell, k)$$

$$c(o) \equiv E(a-\bar{a})\, z(o) = 0.$$

The above programs are sufficient to generate the minimum mean square error recursively:

$$\overline{\epsilon^2(k)} = \overline{\epsilon^2(k-1)} - c_1(k)\, p_1^{-1}(k)\, c_1(k) \qquad k = 1, 2, \ldots n$$

with

$$\epsilon^2(o) = \sigma^2\,.$$

After the measurement $z(k)$ is made, the term $\Delta\, y(k)$ is generated by

$$\Delta y(k) = z(k) - \overline{a^k}\,\overline{x_o} - \sum_{j=1}^{k} m_1(k, k-j)\, p_1^{-1}(k-j)\, \Delta y(k-j)$$

with

$$\Delta y(o) = z(o) - \overline{x_o}\,,$$

and then substituted into

$$\hat{a}(k) = \hat{a}(k-1) + c_1(k)\, p_1^{-1}(k)\, \Delta y(k) \qquad\qquad k = 1,$$

with $\hat{a}(o) = \bar{a}$. This completes the equations needed to generate the FORTRAN program for the linear estimator.

II. QUADRATIC ESTIMATOR

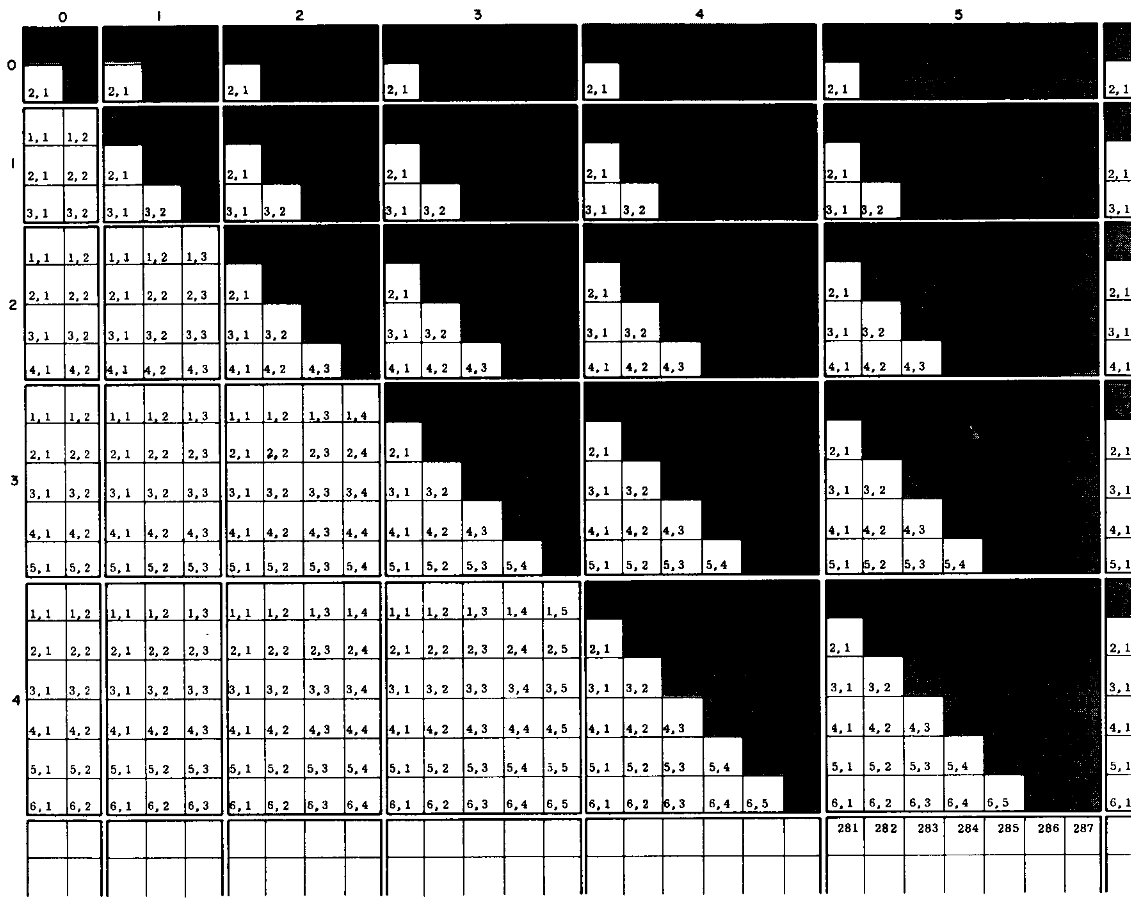
The computer program to generate the optimum quadratic estimate of the plant parameter and its corresponding mean square error were developed by use of the following subprograms:

1. k^{th} moment gaussian. This program was the same as the one used in the linear estimator program (see Part I).
2. Indexing subprogram for mapping the significant terms of the matrix consisting of the submatrices $E\ Y(i)Y^*(j)$ into a linear array. In this way the storage problem is reduced from n^4 to $\frac{n^4}{12}$, an order of magnitude. The significant terms are those terms in the upper triangle of submatrices and the upper triangle of elements of each submatrix, see Figure 4.

   The following equations generate k from α, β, i, j. k is defined only for

   $j = 0, 1, 2, \ldots, n$

   $i = 0, 1, \ldots, j$



ADDRESS OF THE (α, β) ELEMENT OF THE (i, j) SUBMATRIX ASSOCIATED WITH THE MATRIX $[E\,Y_2(i)\,Y^*(j)]$ AND THE CORRESPONDING INDEX NUMBER k.

(α, β) shown in lower left corners, k shown in upper right corners.

FIGURE 4

$1 \leq \alpha \leq i + 2$

$\beta = \alpha,\ \alpha + 1, \ldots, j + 2$

Note however that because of symmetry within the $m_2(i,j)$ matrix, the value of the element at the (α,β) address is the same as the one in the (β,α) address.

Let

$$k_1(j,\alpha) = \begin{cases} 0 & \alpha = 1 \\ \sum\limits_{\ell_1=1}^{\alpha-1} (j+3-\ell_1) & \alpha \geq 2 \end{cases}$$

$$k_2(j,i) = \begin{cases} 0 & i = 0 \\ \sum\limits_{i_2=1}^{i} \sum\limits_{\ell_2=1}^{i_2+1} (j+3-\ell_2) & i \geq 1 \end{cases}$$

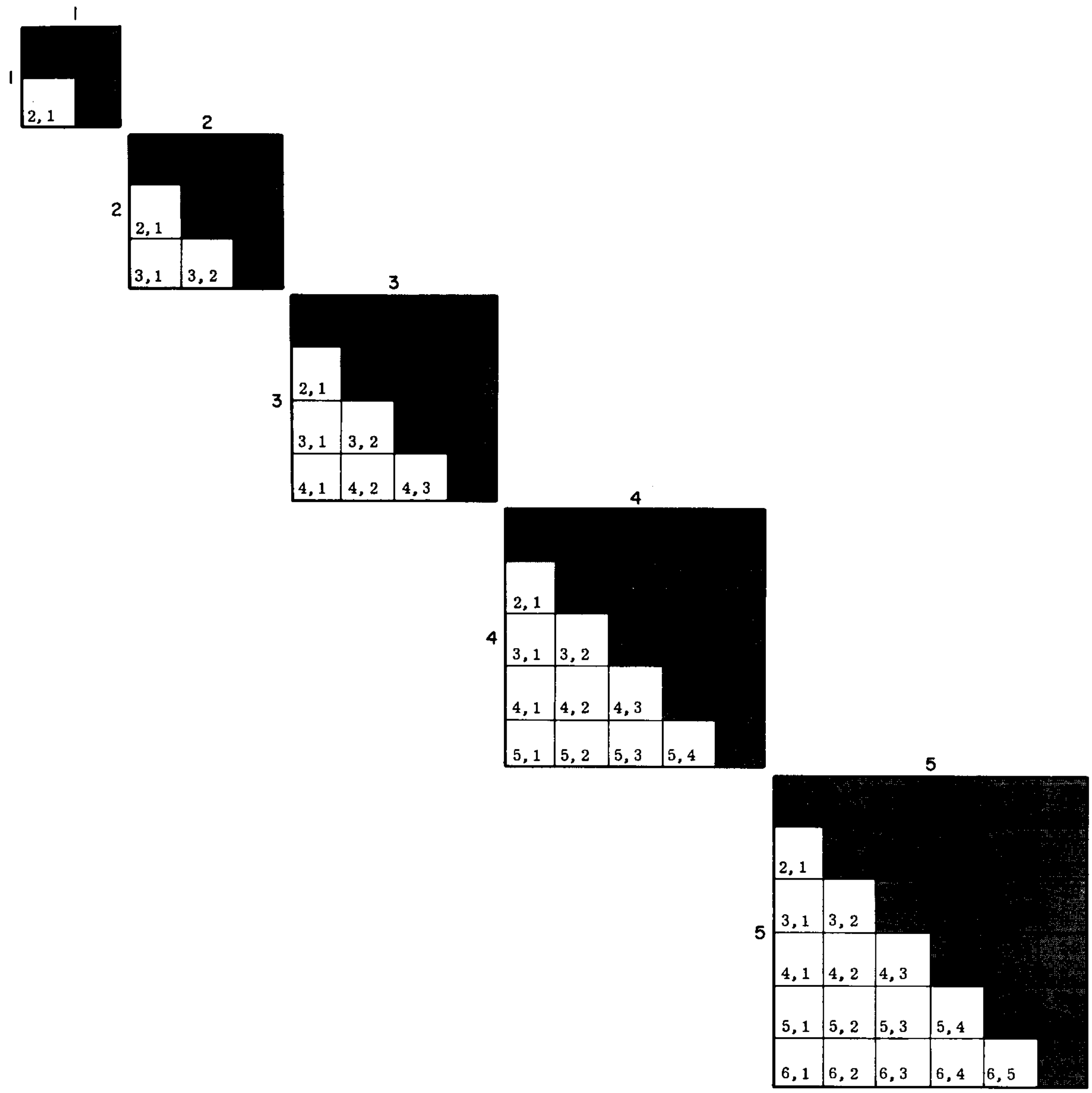
and

$$k_3(j) = \begin{cases} 0 & j = 0 \\ \sum\limits_{j_3=1}^{j} \sum\limits_{i_3=1}^{j_3} \sum\limits_{\ell_3=1}^{i_3+1} (j_3+2-\ell_3) & j \geq 1 \end{cases}$$

then

$$\underline{K} \equiv \underline{K}(j,i;\alpha,\beta) \equiv \begin{cases} k_1(j,\alpha)+k_2(j,i)+k_3(j)+\beta-\alpha+1 & \alpha \leq \beta \\ k_1(j,\beta)+k_2(j,i)+k_3(j)+\alpha-\beta+1 & \alpha \geq \beta \end{cases}$$

3. Indexing subprogram for mapping the significant elements of the $p_2^{-1}(j)$ matrix into a linear array, Figure 5 shows the relationship. An equation which maps these elements in this manner is



ADDRESS OF THE (α, β) ELEMENT OF THE i^{th} MATRIX

$m^{-1}(i, i) = [\, E\, \Delta Y_2(i)\, \Delta Y_2^*(i)\,]^{-1}$ AND ITS INDEX <u>KI</u>.

(α, β) shown in lower left corners, <u>KI</u> in upper right corners.

FIGURE 5

$$\underline{KI}=\underline{KI}(j;\alpha,\beta)=\begin{cases} k_1(j;\alpha)+k_4(j;\alpha,\beta)+k_5(j;\alpha,\beta) & \alpha \le \beta \\ k_1(j;\beta)+k_4(j;\beta,\alpha)+k_5(j;\beta,\alpha) & \alpha \ge \beta \end{cases}$$

where

$$k_1(j;\alpha)=\begin{cases} 0 & \alpha \ge 1, \quad j=0 \\ 0 & \alpha = 1 \quad j \ge 1 \\ \sum_{\ell_1=1}^{\alpha-1} (j+3-\ell_1) & \alpha \ge 2 \quad j \ge 1 \end{cases}$$

$$k_4(j;\alpha,\beta)=\begin{cases} 0 & j=0 \\ \sum_{j_4=1}^{j} \sum_{\ell_4=1}^{j_4+1} (j_4+2-\ell_4) & j \ge 1 \end{cases}$$

$$k_5(j;\alpha,\beta)=\begin{cases} \beta+\alpha-1 & j=0 \\ \beta-\alpha+1 & j \ge 1 \end{cases}$$

The bar below KI is to denote $\underline{KI}$ is the function name of the index in FORTRAN language. That is, for $\alpha \le \beta$,

$$KI=\begin{cases} \beta+\alpha-1 & j=0 \\ \sum_{j_4=1}^{j} \sum_{\ell_4=1}^{j} (j_4-2-\ell_4)+\beta-\alpha+1 & j \geqq 1, \ \alpha=1 \\ \sum_{j_4=1}^{j} \sum_{\ell_4=1}^{j_4+1} (j_4+2-\ell_4)+\beta-\alpha+1+\sum_{\ell_1=1}^{\alpha-1}(j+3-\ell_1) & j \ge 1, \ \alpha \ge 2 \end{cases}$$

and the role of α, β is reversed if $\alpha > \beta$.

4. A ranking routine which takes a set of $\ell = 2, 3,$ or 4 positive integers and rearranges them into a set of integers with increasing magnitudes.

This routine is required in the $E\,z(i_1)\,z(i_2)$, $E\,z(i_1)\,z(i_2)\,z(i_3)$ and $E\,z(i_1)\,z(i_2)\,z(i_3)\,z(i_4)$ subprograms due to the assumptions made that the disturbing noise and the measurement noise be "white".

5. Subprogram to generate $E\,z(i)$ where

$$E\,z(i) = \overline{a^{i}}\;\overline{x_o}$$

6. Subprogram to generate $E\,z(i_1)\,z(i_2)$. For $i_2 \geq i_1$,

$$E\,z(i_2)\,z(i_1) = \overline{a^{i_1+i_2}}\;\overline{x_o^{2}} + R_o \sum_{\ell_1=1}^{i_1} \overline{a^{i_2+i_1-2\ell_1}} + N_o\,\delta_{i_1 i_2}$$

where $\delta_{i_1 i_2}$ is the Kronecker delta function.

7. Subprogram to generate $E\,z(i_3)\,z(i_2)\,z(i_1)$. For $i_3 \geq i_2 \geq i_1$, $E\,z(i_3)\,z(i_2)\,z(i_1) =$

$$= \overline{a^{i_3+i_2+i_1}}\;\overline{x_o^{3}} +$$

$$+\overline{x_o}\,R_o\left[\overline{a^{i_1}} \sum_{\ell_2=1}^{i_2} \overline{a^{i_3+i_2-2\ell_2}} + \overline{a^{i_2}} \sum_{\ell_1=1}^{i_1} \overline{a^{i_3+i_1-2\ell_1}} + \right.$$

$$\left. + \overline{a^{i_3}} \sum_{\ell_1=1}^{i_1} \overline{a^{i_2+i_1-2\ell_1}}\right] + x_o\,N_o\left[\overline{a^{i_1}}\,\delta_{i_2 i_1} + \overline{a^{i_2}}\,\delta_{i_1 i_3} + \overline{a^{i_3}}\,\delta_{i_1 i_2}\right]$$

8. Subprogram to generate $E\,z(i_4)\,z(i_3)\,z(i_2)\,z(i_1)$. For $i_4 \geq i_3 \geq i_2 \geq i_1$,

$$E\, z(i_4)\, z(i_3)\, z(i_2)\, z(i_1) =$$

$$= \overline{a^{i_4+i_3+i_2+i_1}}\,\overline{x_o^4} + \overline{x_o^2}\, R_o \left[3 \sum_{\ell_1=1}^{i_1} \overline{a^{i_4+i_3+i_2+i_1-2\ell_1}} + \right.$$

$$\left. + 2 \sum_{\ell_2=1}^{i_2} \overline{a^{i_4+i_3+i_2+i_1-2\ell_2}} + \sum_{\ell_3=1}^{i_3} \overline{a^{i_4+i_3+i_2+i_1-2\ell_3}} \right]$$

$$+ R_o^2 \left[2 \sum_{\ell_1=1}^{i_1} \sum_{\ell_2=1}^{i_2} \overline{a^{i_4+i_3+i_2+i_1-2\ell_2-2\ell_1}} + \sum_{\ell_1=1}^{i_1} \sum_{\ell_3=1}^{i_3} \overline{a^{i_4+i_3+i_2+i_1-2\ell_3-2\ell_1}} \right]$$

$$+ \overline{x_o^2}\, N_o \left[\overline{a^{i_4+i_3}}\, \delta_{i_2 i_1} + \overline{a^{i_4+i_2}}\, \delta_{i_3 i_1} + \overline{a^{i_4+i_1}}\, \delta_{i_3 i_2} + \overline{a^{i_3+i_2}}\, \delta_{i_4 i_1} + \right.$$

$$\left. + \overline{a^{i_3+i_1}}\, \delta_{i_4 i_2} + \overline{a^{i_2+i_1}}\, \delta_{i_4 i_3} \right] + R_o N_o \left[\sum_{\ell_3=1}^{i_3} \overline{a^{i_4+i_3-2\ell_3}}\, \delta_{i_2 i_1} + \right.$$

$$+ \sum_{\ell_2=1}^{i_2} \overline{a^{i_4+i_2-2\ell_2}}\, \delta_{i_3 i_1} + \sum_{\ell_1=1}^{i_1} \overline{a^{i_4+i_1-2\ell_1}}\, \delta_{i_3 i_2} + \sum_{\ell_2=1}^{i_2} \overline{a^{i_3+i_2-2\ell_2}}\, \delta_{i_4 i_1}$$

$$\left. + \sum_{\ell_1=1}^{i_1} \overline{a^{i_3+i_1-2\ell_1}}\, \delta_{i_4 i_2} + \sum_{\ell_1=1}^{i_1} \overline{a^{i_1+i_2-2\ell_1}}\, \delta_{i_4 i_3} \right]$$

$$+ N_o^2 \left[\delta_{i_4 i_3}\, \delta_{i_2 i_1} + \delta_{i_4 i_2}\, \delta_{i_3 i_1} + \delta_{i_4 i_1}\, \delta_{i_3 i_2} \right]$$

9. Subprogram to generate $\underline{EAY}(\alpha, k)$, i.e.,

$$E(a-\bar{a})\mathcal{Y}(k) \equiv \begin{bmatrix} E(a-\bar{a})\, z_k \\ E(a-\bar{a})\, z_k z_o \\ E(a-\bar{a})\, z_k z_1 \\ \vdots \\ E(a-\bar{a})\, z_k^2 \end{bmatrix} \equiv \begin{bmatrix} \underline{EAY}(1,k) \\ \underline{EAY}(2,k) \\ \underline{EAY}(3,k) \\ \vdots \\ \underline{EAY}(k+2,k) \end{bmatrix}$$

where α is the α^{th} term in the vector $E(a-\bar{a})\mathcal{Y}(k)$. It follows,

$$\underline{EAY}(\alpha, k) = \begin{cases} 0 & k = 0 \\ (\overline{a^{k+1}} - \bar{a}\,\overline{a^k})\,\overline{x_o} & k \geq 1,\ \alpha = 1 \\ (\overline{a^{k+1}} - \bar{a}\,\overline{a^k})\,\overline{x_o^2} & k \geq 1,\ \alpha = 2 \\ \left[\overline{a^{k+\alpha-1}} - \bar{a}\,\overline{a^{k+\alpha-2}}\right]\overline{x_o^2} + & \\ \quad + R_o \sum\limits_{r=1}^{\alpha-2} \left[\overline{a^{k+\alpha-1-2r}} - \bar{a}\,\overline{a^{k+\alpha-2-2r}}\right] & k \geq 1,\ \alpha \geq 3 \end{cases}$$

The bar below EAY is to denote that $\underline{EAY}$ is the name assigned to the FORTRAN function subprogram. The bar below a letter does not denote a vector in Appendix B.

10. Subprogram to invert matrices. The subprogram should also indicate when the determinant is zero and hence, inversion is not permitted. The estimation problem is then singular.

11. Subprogram to generate elements of the (α, β) components of the (i, j) submatrix $E\, Y_2(i)\, Y_2^*(j)$. That is, for a given α, β, i, j such that

$j = 0, 1, 2, \ldots, n$

$i = 0, 1, \ldots, j$

$\alpha = 1, 2, \ldots, i+2$

$\beta = \alpha, \alpha+1, \alpha+2, \ldots, j+2$

$$\underline{E\,Y\,Y} = \begin{cases} E\,z(i)\,z(j) - E\,z(i)\,E\,z(j) & \alpha = 1,\ \beta = 1 \\ E\,z(i)\,z(j)\,z(\beta-2) - E\,z(i)\,E\,z(j)\,z(\beta-2) & \alpha = 1,\ \beta \geq 2 \\ E\,z(i)\,z(\alpha-2)\,z(j)\,z(\beta-2) - \\ \qquad E\,z(i)\,z(\alpha-2)\,E\,z(j)\,z(\beta-2) & \alpha \geq 2,\ \beta \geq 2 \end{cases}$$

If $\alpha > \beta$, α is replaced by β and β by α.

12. Subprogram to generate the following recursion relation:

Let $M_{ij}^{\alpha\beta}$ be the (α, β) component of the $(i, j)^{th}$ submatrix $E\,\Delta\,Y(i)\,Y^*(j)$ and $M_{ji}^{*\alpha,\beta}$ also be the (α, β) component of the $(i, j)^{th}$ submatrix $E\,\Delta\,Y(i)\,Y^*(j)$ for $j \geq i$. Note $E\,\Delta\,Y(i)Y^*(j) = 0$ for $j < i$.

Since $\Delta\,Y(i) = Y(i) - \hat{E}\Big(Y(i)\,\Big|\,Y(i-1), \ldots Y(0)\Big)$.

The (α, β) component of Equation (3.21) is then

$$M_{ij}^{\alpha\beta} = \Big(E\,Y(i)\,Y^*(j)\Big)_{\alpha,\beta}$$

$$- \sum_{\ell=1}^{i} \sum_{\rho_1=1}^{i-\ell+2} \sum_{\rho_2=1}^{i-\ell+2} M_{i,i-\ell}^{*\alpha\,\rho_1} \Big(M_{i-\ell,\,i-\ell}^{-1}\Big)^{\rho_1,\,\rho_2} M_{i-\ell,j}^{\rho_2\beta}$$

Now, if use is made of the indexing subroutines of Steps 2 and 3, the following may be defined

$$\underline{D\,Y\,Y}(\underline{K}) \equiv M_{ij}^{\alpha\beta}$$

where $\underline{K}$ is the index of subprogram 2 and associated with $(j, i; \alpha, \beta)$ and

$$\underline{\mathrm{D\,Y\,YI}}\,(\underline{\mathrm{KI}}) \equiv \left(M^{-1}_{i-\ell,\,i-\ell}\right)^{\rho_1\rho_2}$$

where KI is the index of subprogram 3 and associated with $(i-\ell;\ \rho_1, \rho_2)$. The matrix $M^{-1}_{i-\ell',\,i-\ell'}$ is the matrix $M_{i-\ell',\,i-\ell'}$ inverted. Whereas

$$\left(M^{-1}_{i-\ell',\,i-\ell'}\right)^{\rho_1,\,\rho_2}$$

is the (ρ_1, ρ_2) element of the inverse. The recursion relation for $M^{\alpha\beta}_{ij}$ above becomes

$$\underline{\mathrm{D\,Y\,Y}}(\underline{\mathrm{K}}) = \underline{\mathrm{E\,Y\,Y}}(\underline{\mathrm{K}}) - \sum_{\ell=1}^{i}\ \sum_{\rho_1=1}^{i-\ell+2}\ \sum_{\rho_2=1}^{i-\ell+2} \underline{\mathrm{D\,Y\,Y}}(\underline{\mathrm{KA}}) \bullet$$

$$\bullet\, \underline{\mathrm{D\,Y\,YI}}\,(\underline{\mathrm{KI}}) \bullet \underline{\mathrm{D\,Y\,Y}}\,(\underline{\mathrm{KB}})$$

where K is the index of subprogram 2 generated from $(j, i; \alpha, \beta)$

KA is the index of subprogram 3 generated from $(i, i-\ell; \alpha, \rho_1)$

KI is the index of subprogram 3 generated from $(i-\ell; \rho_1, \rho_2)$

and

KB is the index of subprogram 2 generated from $(j, i-\ell; \rho_2, \beta)$

13. Subprogram to generate elements of the matrix recursion relation, Equation (3.22):

$$C_2^*(j) = E(a-\bar{a})\ Y_1^*(j) - \sum_{\ell=1}^{k} C_2^*(j)\, p_2^{-1}(j-\ell)\ M_2(j-\ell, j)$$

where $C_2(j) = E\, a\, \Delta\, Y_2(j)$. Let $C_2(\alpha; j)$ be the α^{th} element in the vector $C_2(j)$.

Let $\underline{NC} = \frac{(j+1)(j+4)}{2} + \alpha$

$\underline{NC}$ maps the α^{th} element of the j^{th} vector $C_2(j)$ into a linear array. $\underline{NC} = 1, 2, \ldots, \frac{(n+1)(n+4)}{2} + n + 2.$ If $\underline{DC}(\underline{NC})$ denotes the α^{th} element of $C_2(j)$ then

$$\underline{DC}(\underline{NC}) = \underline{EAY}(\alpha;j) - \sum_{\ell=1}^{j-1} \sum_{\rho_1=1}^{j+2-\ell} \sum_{\rho_2=1}^{j+2-\ell} \underline{DC}(\underline{ND}) \bullet$$

$$\bullet \underline{DYYI}(\underline{KI}) \bullet DYY(K)$$

where $\underline{NC} = \frac{(j+1)(j+4)}{2} + \alpha$

$\underline{ND} = \frac{(j-\ell+1)(j-\ell+4)}{2} + \rho$

$\underline{KI}$ is the index associated with the inverse matrix and generated from $(j-\ell;\rho_1, \rho_2)$

$\underline{K}$ is the index generated from $(j, j-\ell;\rho_2, \beta)$

14. Subprogram to generate the associated mean square error:

$$\overline{\epsilon_2^2(j)} = \overline{\epsilon_2^2(j-1)} - \sum_{\rho_1=1}^{j+2} \sum_{\rho_2=1}^{j+2} \underline{DC}(\underline{NC}) \bullet \underline{DYYI}(\underline{KI}) \bullet \underline{DC}(\underline{ND})$$

where

$\underline{NC} = \frac{(j+1)(j+4)}{2} + \rho_1$

$\underline{KI}$ is the index associated with the inverse matrix and generated from $(j;\rho_1, \rho_2)$

$\underline{ND} = \frac{(j+1)(j+4)}{2} + \rho_2$

and

σ^2 is the variance squared of the plant parameter a.

15. Subprogram to generate $\Delta Y_2(k)$ of Equation (3.16), from z(o), z(1), . . . , z(k):

$$\Delta Y_2(k) = Y_2(k) - \sum_{\ell=1}^{k} m_2^*(k, k-\ell) p_2^{-1}(k-\ell) \Delta Y_2(k-\ell)$$

with

$$\Delta Y_2(o) = \begin{bmatrix} z(o) - \overline{z(o)} \\ z^2(o) - \overline{z(o)\, z(o)} \end{bmatrix}$$

The α^{th} component of $Y_2(k)$, denoted $\underline{YA}(k, \alpha)$ is

$$\underline{YA}(k, \alpha) = \begin{cases} z(k) - \overline{z(k)} & \alpha = 1 \\ z(k)\, z(\alpha-2) - \overline{z(k)\, z(\alpha-2)} & \alpha \geq 2 \end{cases}$$

and the α^{th} component of $\Delta Y_2(k)$ denoted $\underline{DY}(k, \alpha)$ is

$$\underline{DY}(k;\alpha) = \underline{YA}(k;\alpha) - \sum_{\ell=1}^{k} \sum_{\rho_1=1}^{k+2-\ell} \sum_{\rho_2=1}^{k+2-\ell} \underline{DYY}(\underline{K}) \bullet$$

$$\bullet\ \underline{DYYI}(\underline{KI}) \bullet DY(\underline{ND})$$

where

$\underline{DYY}(0;\alpha) = \underline{EAY}(0;\alpha)$

$\underline{K}$ is the index of subprogram 2 associated with $(k, k-\ell;\alpha, \rho_1)$

$\underline{KI}$ is the index of subprogram 3 associated with $(k-\ell;\rho_1, \rho_2)$

and

$$\underline{ND} = \frac{(k+1)(k+4)}{2} + \rho_2$$

16. Subprogram to generate the optimum quadratic estimate of the plant parameter a, as given in Chapter III:

$$\hat{a}_2(k) = \hat{a}(k-1) + c_2(k)\, p_2^{-1}(k)\, \Delta Y(k)$$

$$= \hat{a}(k-1) + \sum_{\rho_1=1}^{k+2} \sum_{\rho_2=1}^{k+2} \underline{DC}(\underline{NC}) \quad \underline{DYYI}(\underline{KI})\ \underline{DY}(\underline{ND})$$

where $\hat{a}_2(o) = \bar{a}$,

$$\underline{NC} = \frac{(k+1)(k+4)}{2} + \rho_1 ,$$

$\underline{KI}$ is the index of subprogram 3 associated with $(k; \rho_1, \rho_2)$,

and

$$ND = \frac{(k+1)(k+4)}{2} + \rho_2 ,$$

This completes the description of the computer program for the quadratic estimator.

APPENDIX C

THEOREM 6.1 AND PROOF

Theorem 6.1 (repeated for convenience)

If the conditions of Section 6.1 are satisfied, then

(i) $$W^{(o)} = E\,a - \int_0^T\int_0^T W^{(2)}(t_1, t_2)\,E\,z(t_1)\,z(t_2)\,dt_1\,dt_2 \qquad (6.3)$$

(ii) $$W^{(1)}(t) = 0 \qquad (6.4)$$

(iii) $W^{(2)}(t_1, t_2)$ satisfies the integral equation

$$W^{(2)}(t_3, t_4) =$$

$$\frac{\lambda^2}{2} M_{2o}(t_3, t_4) - \frac{\lambda^2}{2}\int_0^T\int_0^T W^{(2)}(t_1, t_2)\,g_2(t_1, t_2; t_3, t_4)\,dt_1\,dt_2$$

$$- \frac{\lambda}{2}\left\{\int_0^T W^{(2)}(t_1, t_4)\,g_1(t_1, t_3)\,dt_1 + \int_0^T W^{(2)}(t_1, t_3)\,g_1(t_1, t_4)\,dt_1\right.$$

$$\left. + \int_0^T W^{(2)}(t_4, t_2)\,g(t_2, t_3)\,dt_2 + \int_0^T W^{(2)}(t_3, t_2)\,g_1(t_2, t_4)\,dt_2\right\} \qquad (6.5)$$

where

$$M_{2o}(t_1, t_2) = E(a-m)\,z(t_1)\,z(t_2)$$

$$g_1(t_1, t_2) = V_o^2\,E\,e^{a(t_1+t_2)}$$

$$g_2(t_1, t_2; t_3, t_4) = 3\,V_o^4\,E\,e^{a(t_1+t_2+t_3+t_4)} - V_o^4\,E\,e^{a(t_1+t_2)}\,E\,e^{a(t_3+t_4)}$$

(iv) The quadratic estimator is

$$\hat{a}_2 = E\,a + \int_0^T\int_0^T W^{(2)}(t_1, t_2)\left[z(t_1)\,z(t_2) - E\,z(t_1)\,z(t_2)\right]dt_1\,dt_2 \qquad (6.6)$$

(v) The minimum mean square error is

$$e_2^2 = \sigma^2 - \int_0^T\int_0^T W^{(2)}(t_1, t_2) V_o^2\, E(a-\bar{a})\, e^{a(t_1+t_2)} dt_1\, dt_2 \qquad (6.7)$$

Proof to Theorem 6.1

Substitute Equation (6.1) into Equation (6.2). There results directly

$$W^{(o)} = \bar{a} - \int_0^T W^{(1)}(t_1)\, E\, z(t_1)\, dt_1 - \int_0^T\int_0^T W^{(2)}(t_1, t_2)\, E\, z(t_1)\, z(t_2)\, dt_1\, dt_2 \qquad (C.1)$$

Q. E. D. Part (i)

and substituting this equation too into Equation (6.2), gives the following two integral equations

$$M_{1o}(t_3) = \int_0^T W^{(1)}(t_1)\, M_{11}(t_1, t_3)\, dt_1 + \int_0^T\int_0^T W^{(2)}(t_1, t_2)\, M_{12}(t_1, t_2; t_3)\, dt_1\, dt_2 \qquad (C.2)$$

$$M_{2o}(t_3, t_4) = \int_0^T W^{(1)}(t_1)\, M_{21}(t_1; t_3, t_4)\, dt_1 + \int_0^T\int_0^T W^{(2)}(t_1, t_2)\, M_{22}(t_1, t_2; t_3, t_4)\, dt_3\, dt_4 \qquad (C.3)$$

where, since $\bar{x}_o = 0$,

$$M_{1o}(t_3) = E(a-\bar{a})\, z(t_3) = 0 \qquad (C.4)$$

$$M_{11}(t_1, t_3) = E\, z(t_1)\, z(t_3) - E\, z(t_1) E\, z(t_2) = V_o^2 E\, e^{a(t_1+t_3)} + N_o \ (t_1 - t_2) \qquad (C.5)$$

$$M_{12}(t_1; t_3, t_4) = E\, z(t_1)\, z(t_3)\, z(t_4) - E\, z(t_1)\, E\, z(t_3)\, z(t_4) = 0 \qquad (C.6)$$

$$M_{2o}(t_3, t_4) = E(a-\bar{a})\, z(t_3)\, z(t_4) = V_o^2 E\left[(a-\bar{a})\, e^{a(t_1+t_2)}\right] \tag{C.7}$$

$$M_{21}(t_1, t_2; t_4) = E\, z(t_1)\, z(t_2)\, z(t_4) - E\, z(t_1)\, z(t_2)\, E\, z(t_4) = 0 \tag{C.8}$$

$$\begin{aligned} M_{22}(t_1, t_2; t_3, t_4) &= E\, z(t_1)\, z(t_2)\, z(t_3)\, z(t_4) - \\ &- E\, z(t_1)\, z(t_2)\, E\, z(t_3)\, z(t_4) \\ &= 3\, V_o^4\, E\, e^{a(t_1+t_2+t_3+t_4)} - V_o^4\, E\, e^{a(t_1+t_2)}\, E\, e^{a(t_3+t_4)} \\ &+ V_o^2 N_o E \left\{ e^{a(t_1+t_3)} \delta_{24} + e^{a(t_1+t_4)} \delta_{23} + e^{a(t_2+t_3)} \delta_{14} + \right. \\ &\left. + e^{a(t_2+t_4)} \delta_{13} \right\} + N_o^2 \left[\delta_{13}\, \delta_{14} + \delta_{14}\, \delta_{23}\right] \end{aligned} \tag{C.9}$$

where $\delta_{ij} \equiv$ impulse function $\equiv \delta(t_i - t_j)$. Substituting Equations (C.4) to (C.6) into Equation (C.2) yields

$$\int_0^T W^{(1)}(t_1) \left[V_o^2 E\, e^{a(t_1+t_3)} + N_o\, \delta\,(t_1 - t_2)\right] dt_1 = 0 \tag{C.10}$$

A sufficient condition for this to hold is for $W^{(1)}(t_1) \equiv 0$. Therefore set

$$W^{(1)}(t) \equiv 0 \qquad\qquad 0 \le t \le T \tag{C.11}$$

This proves Part (iii).

Equation (6.4) in Equation (C.3) yields

$$\begin{aligned} M_{2o}(t_3, t_4) = &\left[\left[\int_0^T\!\!\int_0^T W^{(2)}(t_1, t_2)\, g_{oo}(t_1, t_2; t_3, t_4)\, dt_1\, dt_2 + \right.\right. \\ &+ V_o^2 N_o \left\{ \int_0^T W^{(2)}(t_1, t_4)\, E\, e^{a(t_1+t_3)} dt_1 + \int_0^T W^{(2)}(t_1, t_3)\, E\, e^{a(t_1+t_4)} dt_1 + \right. \\ &\left.\left.\left. + \int_0^T W^{(2)}(t_4, t_2)\, E\, e^{a(t_2+t_3)} dt_2 + \int_0^T W^{(2)}(t_3, t_2)\, E\, e^{a(t_2+t_4)} dt_2 \right\} \right]\right] + \\ &+ N_o^2 \left[W^{(2)}(t_3, t_4) + W^{(2)}(t_4, t_3)\right] \end{aligned} \tag{C.12}$$

where

$$g_{oo}(t_1,t_2;t_3,t_4) = 3\,V_o^4\,E\,e^{a(t_1+t_2+t_3+t_4)} - V_o^4\,E\,e^{a(t_1+t_2)}\,E\,e^{a(t_3+t_4)} \quad (C.13)$$

Equation (C.12) is a Fredholm equation of the second kind in two variables. From the symmetry of Equation (C.12) in (t_3, t_4) and (t_4, t_3) we can without loss of generality assume $W(t_3, t_4)$ symmetric. Let $\lambda = \frac{1}{N_o}$. Then using the assumption of symmetry $W(t_3, t_4) = W(t_4, t_3)$, and rearranging

$$W^{(2)}(t_3,t_4) = \frac{\lambda^2}{2} M_{2o}(t_3,t_4) - \frac{\lambda^2}{2}\left[\!\left[\qquad \right]\!\right]$$

where the double bracketed term denotes the collection of terms inside the double brackets of Equation (C.12). The $W^{(2)}(t_3, t_4)$ which satisfies Equation (C.12) gives the desired quadratic weighting function. Substitute Equations (C.1) and (C.11) into Equation (6.1). There results

$$\hat{a}_2 = E\,a + \int_0^T\!\!\int_0^T W^{(2)}(t_1,t_2)\left[z(t_1)\,z(t_2) - E\,z(t_1)\,z(t_2)\right] dt_1\,dt_2$$

where $W^{(2)}(t_1, t_2)$ satisfies Equation (C.14). The desired expression for the continuous quadratic estimator has been developed.

Q.E.D. Part (iv)

The mean square error of the quadratic estimator

$\overline{e_2^2} = E(a-\hat{a}_2)^2$ becomes

$$\overline{e_2^2} = \sigma^2 - \int_0^T\!\!\int_0^T W^{(2)}(t_1,t_2)\,E\left\{a\left[z(t_1)\,z(t_2) - E\,z(t_1)\,z(t_2)\right]\right\} dt_1\,dt_2$$

or

$$e_2^2 = \sigma^2 - \int_0^T \int_0^T W^{(2)}(t_1, t_2)\, V_o^2\, E\left[(a-\bar{a})\, e^{a(t_1 + t_2)}\right] dt_1\, dt_2$$

Q.E.D. Theorem 6.1